This is a repository copy of *Uniform Consistency of Nonstationary Kernel-Weighted Sample Covariances for Nonparametric Regression*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/102283/

Version: Accepted Version

Article:

Li, Degui orcid.org/0000-0001-6802-308X, Phillips, Peter C. B. and Gao, Jiti (2016) Uniform Consistency of Nonstationary Kernel-Weighted Sample Covariances for Nonparametric Regression. Econometric Theory. pp. 655-685. ISSN 0266-4666

https://doi.org/10.1017/S0266466615000109

Reuse
["licenses_typename_other" not defined]

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Uniform Consistency of Nonstationary Kernel-Weighted Sample Covariances for Nonparametric Regression

Degui Li[*], Peter C. B. Phillips[†] and Jiti Gao[‡]

December 18, 2014

Abstract

We obtain uniform consistency results for kernel-weighted sample covariances in a nonstationary multiple regression framework that allows for both fixed design and random design coefficient variation. In the fixed design case these nonparametric sample covariances have different uniform asymptotic rates depending on direction, a result that differs fundamentally from the random design and stationary cases. The uniform asymptotic rates derived exceed the corresponding rates in the stationary case and confirm the existence of uniform super-consistency. The modelling framework and convergence rates allow for endogeneity and thus broaden the practical econometric import of these results. As a specific application, we establish uniform consistency of nonparametric kernel estimators of the coefficient functions in nonlinear cointegration models with time varying coefficients or functional coefficients, and provide sharp convergence rates. For the fixed design models, in particular, there are two uniform convergence rates that apply in two different directions, both rates exceeding the usual rate in the stationary case.

Key words and phrases: Cointegration; Functional coefficients; Kernel degeneracy; Nonparametric kernel smoothing; Random coordinate rotation; Super-consistency; Uniform convergence rates; Time varying coefficients.

JEL classification: C13, C14, C32.

[*]University of York

[†]Yale University, University of Auckland, Southampton University, and Singapore Management University

[‡]Monash University

1 Introduction

Uniform consistency results with convergence rates for nonparametric kernel estimators have been extensively studied in the existing literature. These results are important in many kernel-based applications such as semiparametric estimation with first-stage kernel smoothing, kernel-based specification testing, and cross-validation bandwidth selection. Existing studies mainly focus on obtaining uniform consistency results for independent and identically distributed (*i.i.d.*) data or time series that satisfy certain stationarity and mixing conditions. Early statistical studies include Mack and Silverman (1982), Roussas (1990), Liebscher (1996), Masry (1996) and Bosq (1998). Later developments and econometric applications can be found in Hansen (2008), Kristensen (2009) and Li *et al* (2012).

Recent years have witnessed a growing literature on nonparametric kernel smoothing in a nonstationary framework. This work is of practical importance because the stationarity condition is restrictive and unrealistic in many empirical applications as discussed in the literature. Among others, see Phillips and Park (1998), Karlsen and Tjøstheim (2001), Karlsen *et al* (2007), Cai *et al* (2009), Wang and Phillips (2009a, 2009b), Xiao (2009), Chen *et al* (2010), Chen *et al* (2012), and Gao and Phillips (2013a, 2013b). Most recently, there has been interest in obtaining uniform consistency results for nonparametric kernel smoothing under nonstationarity (notably, Chan and Wang, 2012; Duffy, 2013; Wang and Wang, 2013; Wang and Chan, 2014; Gao *et al,* 2015). This work confirms that uniform convergence rates of kernel-based estimates in nonstationary cases are slower than those in the stationary case. Just as in pointwise convergence, the slower convergence rate is explained by the random wandering character of nonstationary time series (such as those arising in unit root or null recurrent Markov frameworks) so that the amount of time spent by the series in the vicinity of any particular point is of smaller order than the stationary case, thereby reducing the effective sample size in estimation.

This paper develops uniform consistency results for potentially multivariate kernel-weighted sample covariance functions of the following form

$$Q_n(z) = \sum_{t=1}^{n} K\left(\frac{Z_t - z}{h}\right) X_t e_t \tag{1.1}$$

after appropriate normalization, where $K(\cdot)$ is a kernel function, $h \equiv h_n$ is a bandwidth which tends to zero as n tends to infinity, X_t is a nonstationary I(1) process with dimension

$d \geq 1$, and e_t is stationary. Detailed properties of the time series X_t and e_t are provided in Section 2. Quantities such as the weighted sample covariance (1.1) play a central role in kernel regression and are fundamental in determining the limit theory of such regressions. For example, when X_t and e_t are stationary and Z_t follows an *i.i.d.* random design, the standardized quantity $Q_n(z)/\big[\sum_{t=1}^n K(\frac{Z_t-z}{h})\big]$ estimates the conditional covariance between X_t and e_t for given $Z_t = z$. When X_t is integrated, a similar quantity arises in nonparametric functional coefficient cointegrating regression. Section 4 shows that the limit theory for $Q_n(z)$ is useful in deriving asymptotics for kernel estimation of coefficient functions in nonlinear cointegration models with varying coefficients. This paper focuses on two cases of particular interest: (i) $Z_t = \frac{t}{n}$, corresponding to a fixed design structure; and (ii) Z_t is *i.i.d.*, corresponding to a random design framework. Asymptotic rates for $Q_n(z)$ will be developed uniformly in z for these two cases.

For case (ii) with random design Z_t, we show that the uniform asymptotic rate of (1.1) is $O_P(n\sqrt{h \log n})$, which exceeds the $O_P(\sqrt{nh \log n})$ rate that holds when both X_t and e_t are stationary. This result can be used to derive a uniform convergence rate for nonparametric kernel-based estimation of the functional coefficients in nonlinear cointegration models where super-consistency exists. In contrast, case (i) with fixed design Z_t is much more complicated because kernel weighting produces degeneracy in the signal matrix defined later on the left hand side of (2.3) when X_t has dimension $d > 1$. This degeneracy introduces a major challenge in developing the asymptotic estimation theory as shown in other recent work (Phillips *et al*, 2013). This phenomenon of *kernel degenerate asymptotics* is new and will be discussed in Section 2 where we show how the limit theory may be developed to accommodate the degeneracy. Theorem 2.1 and Corollary 2.1 below show that the uniform rates for the quantity $Q_n(z)$ depends on a certain random direction, yielding two different rates both of which exceed the $O_P(\sqrt{nh \log n})$ rate that applies in the stationary case.

These results are used to derive uniform consistency for nonparametric kernel estimates in nonlinear cointegration models with varying coefficients, accommodating the super-consistency rates of kernel convergence. Our approach allows for endogeneity between the regressor X_t and the error e_t, which enhances the practical relevance of the results in cointegration analysis: case (i) with the fixed design framework $Z_t = \frac{t}{n}$ relates particularly to cointegration models with time-varying coefficients (Park and Hahn, 1999; Phillips *et al*, 2013); and case (ii) with random design Z_t relates to cointegration models with functional coefficients (Cai *et al*, 2009; Xiao, 2009; Gao and Phillips 2013b). In addition, the uniform consistency results with sharp convergence rates that are obtained in this paper are of some independent interest with other potential applications, such as to semiparametric cointegration models with partially-varying coefficients.

The remainder of the paper is organised as follows. Uniform consistency results for the fixed design case are given in Section 2. Those for the random design case are given in Section 3. Applications of the main results to nonlinear cointegration models with varying coefficients are provided in Section 4. Section 5 concludes the paper. Proofs of the main results are given in the Appendix.

2 Uniform rates with a fixed design covariate

This section establishes uniform asymptotic rates for $Q_n(z)$ defined in (1.1) with $Z_t = \frac{t}{n}$. The random design case is discussed in Section 3. We start with regularity conditions that characterize the multivariate nonstationary time series X_t and the scalar stationary process e_t. Let X_t be a unit root process with generating mechanism $X_t = X_{t-1} + v_t$, initial value $X_0 = O_P(1)$ and innovations determined by the linear process

$$v_t = \mathbf{\Phi}(\mathcal{L})\varepsilon_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \varepsilon_{t-j}, \tag{2.1}$$

where $\mathbf{\Phi}(\mathcal{L}) = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \mathcal{L}^j$, $\mathbf{\Phi}_j$ is a sequence of $d \times d$ matrices, $\mathcal{L}$ is the lag operator and $\{\varepsilon_t\}$ is a sequence of *i.i.d.* innovation vectors with dimension d.

Assumption 1. *(i) Let $\{\varepsilon_t\}$ be i.i.d. d-dimensional random vectors with $\mathbb{E}[\varepsilon_t] = 0$, $\mathbf{\Lambda}_\varepsilon \equiv \mathbb{E}\left[\varepsilon_t \varepsilon_t'\right]$ positive definite, and $\mathbb{E}\left[\|\varepsilon_t\|^{4+\delta_0}\right] < \infty$ for $\delta_0 > 0$. The linear process coefficient matrices in (2.1) satisfy that $\sum_{j=0}^{\infty} j\|\mathbf{\Phi}_j\| < \infty$ and $\mathbf{\Omega}_\varepsilon \equiv \mathbf{\Phi}\mathbf{\Lambda}_\varepsilon\mathbf{\Phi}'$ is positive definite with $\mathbf{\Phi} = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \neq 0$, where $\|\cdot\|$ denotes the Euclidean norm.*

(ii) Let $\{e_t\}$ be generated by the linear process $e_t = \sum_{j=0}^{\infty} \phi_j \eta_{t-j}$, where $\{\eta_t\}$ is an i.i.d. sequence with $\mathbb{E}[\eta_t] = 0$, $\sigma_\eta^2 \equiv \mathbb{E}[\eta_t^2] > 0$, $\mathbb{E}\left[|\eta_t|^{4+\delta_0}\right] < \infty$, $\phi \equiv \sum_{j=0}^{\infty} \phi_j \neq 0$, and $\sum_{j=0}^{\infty} j|\phi_j| < \infty$. In addition, (η_t, ε_t') is independent of $\{(\eta_s, \varepsilon_s') : s \leq t-1\}$, but η_t may be correlated with ε_t.

Assumption 1(i) ensures that a functional law holds for X_t upon standardization. In particular, from Phillips and Solo (1992) we have for $t = \lfloor nx \rfloor$ and $0 < x \le 1$,

$$\frac{X_t}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{s=1}^{t} v_s + \frac{1}{\sqrt{n}}X_0 = \frac{1}{\sqrt{n}}\sum_{s=1}^{\lfloor nx \rfloor} v_s + o_P(1) \Rightarrow B_x(\mathbf{\Omega}_\varepsilon), \tag{2.2}$$

where $B_\cdot(\mathbf{\Omega}_\varepsilon)$ is a d-dimensional Brownian motion with variance matrix $\mathbf{\Omega}_\varepsilon$ and the floor function $\lfloor \cdot \rfloor$ denotes integer part. In a more specialized setting, Assumption 1(ii) might be replaced by a martingale difference structure with $\mathbb{E}\left[e_t|\mathcal{G}_{t-1}\right] = 0$ *a.s.*, where $\mathcal{G}_t = \sigma(e_t, \cdots, e_1, \varepsilon_{t+1}, \varepsilon_t, \cdots)$, and the uniform consistency results developed in this paper still hold. Instead, we allow for a more general linear dependence structure and joint contemporaneous correlation between the innovations η_t and ε_t which builds endogeneity into the regression equation. Moreover, the limit theory developed in this paper continues to hold with some modification of the proofs when e_t and v_t are jointly determined by a multivariate linear process of the form

$$(e_t, v_t')' = \mathbf{\Phi}^*(\mathcal{L})\varepsilon_t^* = \sum_{j=0}^{\infty} \mathbf{\Phi}_j^* \varepsilon_{t-j}^*,$$

where $\mathbf{\Phi}^*(\mathcal{L}) = \sum_{j=0}^{\infty} \mathbf{\Phi}_j^* \mathcal{L}^j$ with $\mathbf{\Phi}_j^*$ a sequence of $(d+1) \times (d+1)$ coefficient matrices and $\{\varepsilon_t^*\}$ is a sequence of *i.i.d.* random vectors of dimension $d+1$.

We impose some mild conditions on the kernel function $K(\cdot)$ and the bandwidth h.

Assumption 2. *(i) The kernel function $K(\cdot)$ is continuous, positive, symmetric and has compact support $[-1, 1]$ with $\mu_0 \equiv \int_{-1}^{1} K(u)du = 1$.*

(ii) The bandwidth h satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

A recent paper by Phillips *et al* (2013) shows that for $0 < z \le 1$,

$$\frac{1}{n^2 h}\sum_{t=1}^{n} X_t X_t' K\left(\frac{t - nz}{nh}\right) \Rightarrow \mathbf{W}_z(\mathbf{\Omega}_\varepsilon), \tag{2.3}$$

where $\mathbf{W}_z(\mathbf{\Omega}_\varepsilon) = B_z(\mathbf{\Omega}_\varepsilon)B_z(\mathbf{\Omega}_\varepsilon)'$ and "$\Rightarrow$" denotes weak convergence. However, the $d \times d$ limiting Wishart matrix $\mathbf{W}_z(\mathbf{\Omega}_\varepsilon)$ on the right hand side of (2.3) is an outer product of the Gaussian variate $B_z(\mathbf{\Omega}_\varepsilon)$ and hence is singular with rank unity when $d > 1$. In other words, the limit Wishart variate $\mathbf{W}_z(\mathbf{\Omega}_\varepsilon)$ has a single degree of freedom and is a singular distribution when $d > 1$. It follows from (2.3) that the kernel-weighted signal matrix $(1/n^2 h)\sum_{t=1}^{n} X_t X_t' K\left(\frac{t-nz}{nh}\right)$ is asymptotically singular whenever the dimension of the regressor X_t exceeds unity. This phenomenon of kernel degeneracy leads to asymptotic singularity in the limit distribution and variance matrix of the corresponding kernel-weighted sample covariance $\frac{1}{n\sqrt{h}}Q_n(z)$ when Z_t is a fixed design structure.

The reason for this kernel degeneracy in the limit of the weighted signal matrix is that kernel regression concentrates attention on some time coordinate (say z_0), thereby fixing attention on a particular coordinate of the limit process of the regressor, say $X_{\lfloor nz_0 \rfloor}$. In the multivariate case with $d > 1$, this focus on a single time coordinate produces a limit signal matrix (corresponding to the limit of the outer product $\frac{1}{n}X_{\lfloor nz_0 \rfloor}X'_{\lfloor nz_0 \rfloor}$) that is of deficient rank unity. Moreover, the zero eigenspace of this limit matrix depends on the (random vector) value of the limit process at that time coordinate. To address this type of kernel degeneracy, Phillips *et al* (2013) develop a coordinate transformation to isolate the random direction of singularity and use the associated coordinate rotation to obtain the limit distribution theory. We extend this technique in the present paper to derive uniform asymptotic rates for $Q_n(z)$. For $z > h$, define the quantities $\gamma_n(z) = \lfloor n(z-h) \rfloor$,

$$q_{\gamma_n(z)} = \frac{b_{\gamma_n(z)}}{\left[b'_{\gamma_n(z)}b_{\gamma_n(z)}\right]^{1/2}} = \frac{b_{\gamma_n(z)}}{\|b_{\gamma_n(z)}\|}, \text{ and } b_{\gamma_n(z)} = \frac{1}{\sqrt{n}}X_{\gamma_n(z)}.$$

Let $q^{\perp}_{\gamma_n(z)}$ be an orthogonal complement of $q_{\gamma_n(z)}$ constructed so that

$$\mathbf{D}_n(z)'\mathbf{D}_n(z) = \mathbf{I}_d \quad \text{with} \quad \mathbf{D}_n(z) = \left[q_{\gamma_n(z)}, q^{\perp}_{\gamma_n(z)}\right], \tag{2.4}$$

and introduce the normalization matrix

$$\mathbf{R}_n = \text{diag}\left\{n\sqrt{h},\ (nh)\mathbf{I}_{d-1}\right\}, \tag{2.5}$$

where $\mathbf{I}_r$ is the $r \times r$ identity matrix. The transformation matrix $\mathbf{D}_n(z)$ is random, path dependent, and localized to the coordinate of concentration at $\gamma_n(z)$. Within this transformation matrix, the component vector $q_{\gamma_n(z)}$ and complementary submatrix $q^{\perp}_{\gamma_n(z)}$ provide random directions localized according to $\gamma_n(z)$.

The following result gives uniform asymptotic orders for $Q_n(z)$.

THEOREM 2.1. *Suppose that Assumptions 1 and 2 are satisfied. Let*

$$\frac{n^{\delta_0-4}h^{\delta_0+12}}{(\log n)^{\delta_0}} \to \infty, \tag{2.6}$$

where $\delta_0 > 4$ is defined as in Assumption 1(i). Then, we have

$$\sup_{\epsilon_* \le z \le 1-\epsilon_*} \left\|\mathbf{R}_n^{-1}\mathbf{D}_n(z)'Q_n(z)\right\| = O_P(\sqrt{\log n}), \tag{2.7}$$

where $\epsilon_* > 0$ can be arbitrarily small.

In the limit result (2.7), we take the suprema over the closed interval $[\epsilon_*, 1-\epsilon_*]$ and the upper limit can be extended from $1-\epsilon_*$ to 1 with minor modification of the proofs. However, we may not extend the lower limit from ϵ_* to 0. For example, when $z = h$, we have $\gamma_n(z) = 0$ which indicates that $b_{\gamma_n(z)} = 0$ and thus $q_{\gamma_n(z)}$ is undefined. The condition (2.6) indicates the trade-off between the moment condition and bandwidth restriction. In particular, when both η_t and ε_t are Gaussian, the value of δ_0 can be arbitrarily large, the condition (2.6) would be close to the commonly-used one of $nh \to \infty$. From the above theorem, we find that the existence of correlation between X_t and e_t does not affect the uniform rate of the kernel-weighted sample covariance. This robustness to endogeneity in the present case arises because the induced asymptotic bias arising from the non-zero mean of $Q_n(z)$ turns out to be a "second order" bias effect as in the linear parametric case (Phillips and Durlauf, 1986; Phillips and Hansen, 1990). Furthermore, from the definitions of $\mathbf{D}_n(z)$ and $\mathbf{R}_n$, it is apparent that two different rates are obtained for the two directions determined by $q_{\gamma_n(z)}$ and $q^{\perp}_{\gamma_n(z)}$, which are stated in the following corollary.

Corollary 2.1. *Let the assumptions in Theorem 2.1 hold. Then, we have*

$$\sup_{\epsilon_* \le z \le 1-\epsilon_*} \left| q'_{\gamma_n(z)} Q_n(z) \right| = O_P(n\sqrt{h \log n}) \tag{2.8}$$

and

$$\sup_{\epsilon_* \le z \le 1-\epsilon_*} \left\| (q^{\perp}_{\gamma_n(z)})' Q_n(z) \right\| = O_P(nh\sqrt{\log n}). \tag{2.9}$$

The above results are used in Section 4 to derive uniform convergence rates for nonparametric kernel-based estimators of the time-varying coefficients in nonlinear cointegration models. Although the uniform rates are different in the two directions, both rates exceed the usual rate $O_P\left(\sqrt{nh \log n}\right)$ for kernel estimators that applies in stationary models. A detailed discussion of this phenomenon in the point-wise kernel regression case is given in Phillips *et al* (2013). We provide the following example to illustrate the above results.

Example 2.1. Let $X_t = (X_{1t}, X_{2t})'$, $X_{it} = X_{i,t-1} + v_{it}$ for $i = 1$ and 2, $v_{it} = \rho_i v_{i,t-1} + \varepsilon_{it}$, and $e_t = \rho e_{t-1} + \eta_t$, where $(\varepsilon_t, \eta_t) \equiv (\varepsilon_{1t}, \varepsilon_{2t}, \eta)' \sim_{iid} N(0, V)$ with $V > 0$. If $-1 < \rho_i, \rho < 1$, it is easy to verify that Assumption 1 above is satisfied. With this model structure, we can write out the specific form of the transformation matrix, from which the two random directions can be determined. Let $\gamma_n(z)$ be defined as before,

$$q_{\gamma_n(z)} = [\frac{X_{1\gamma_n(z)}}{\sqrt{n}\|b_{\gamma_n(z)}\|}, \frac{X_{2\gamma_n(z)}}{\sqrt{n}\|b_{\gamma_n(z)}\|}]' \equiv \big[q_{1n}(z), q_{2n}(z)\big]' \text{ with } b_{\gamma_n(z)} = \frac{1}{\sqrt{n}} X_{\gamma_n(z)}$$

and

$$q^{\perp}_{\gamma_n(z)} = \big[p_{1n}(z), p_{2n}(z)\big]' = \big[q_{2n}(z), -q_{1n}(z)\big]' \text{ or } q^{\perp}_{\gamma_n(z)} = \big[-q_{2n}(z), q_{1n}(z)\big]',$$

so that the orthogonality condition in (2.4) is satisfied. By Theorem 2.1, when Z_t has fixed design, the uniform rate of $Q_n(z)$ is $O_P(n\sqrt{h\log n})$ in the direction determined by $q_{\gamma_n(z)}$, whereas the uniform rate of $Q_n(z)$ is $O_P(nh\sqrt{\log n})$ in the direction orthogonal to $q_{\gamma_n(z)}$. Both directions are random and are determined by $\gamma_n(z) = \lfloor n(z-h)\rfloor$. The uniform rate $O_P(n\sqrt{h\log n})$ can be understood as $O_P(\sqrt{n^2 h \log n})$, which indicates that the effective sample size used to derive the uniform rate of $Q_n(z)$ in the direction $q_{\gamma_n(z)}$ is of order (n^2h). In contrast, the effective sample size used to derive the uniform rate of $Q_n(z)$ in the direction $q^{\perp}_{\gamma_n(z)}$ is of order $(nh)^2$, which is smaller than that in the direction $q_{\gamma_n(z)}$ and thus leads to a smaller uniform rate for $Q_n(z)$ in the direction orthogonal to $q_{\gamma_n(z)}$. Hence, the signal (and convergence rate) of the kernel weighted sample covariance $Q_n(z)$ is strongest in the direction $q_{\gamma_n(z)}$, which spans the direction of the dominating range space of the (asymptotically signal) signal matrix $(1/n^2h)\sum_{t=1}^n X_t X_t' K\left(\frac{t-nz}{nh}\right)$, and this rate is uniformly $O_P(\sqrt{n^2 h\log n})$.

3 Uniform rates with a random design covariate

This section develops uniform asymptotic rates for the sample covariance $Q_n(z)$ when Z_t is generated by *i.i.d.* random variables, and compares this result with those of the fixed design case studied in the previous section. For the stationary case, it is well known that the same uniform rates hold for $Q_n(z)$ irrespective of whether Z_t is a random design or fixed design variate. In contrast to Section 2, there is no kernel degeneracy in the random design case and a common uniform convergence rate applies which is the same as that given in (2.8). The next assumption is used in the derivation of the uniform consistency result in Theorem 3.1 below.

Assumption 3. *Let $\{(Z_t, \eta_t, \varepsilon_t')\}$ be a sequence of i.i.d. random vectors with continuous density function $f(\cdot,\cdot,\cdot)$, and let Z_t be independent of η_t and have compact support, say* $[0,1]$.

Much of the existing literature on the limit theory of $Q_n(\cdot)$ for the random design case imposes a martingale difference structure on e_t, which excludes the possibility of correlation between X_t and e_t (c.f., Cai *et al*, 2009; Li *et al*, 2014). However, for consistency with the framework of Section 2, we follow the same structure as Assumption 1 to generate the unit root process X_t and the stationary process e_t, thereby allowing for possible correlation between X_t and e_t. Hence, the result below has wider applicability than currently available theory.

The uniform asymptotic order for $Q_n(z)$ in the random design case is given as follows.

THEOREM 3.1. *Suppose that Assumptions 1–3 are satisfied. Let*

$$\frac{n^{2+\delta_0}h^{4+\delta_0}}{(\log n)^{4+\delta_0}} \to \infty, \tag{3.1}$$

where δ_0 is defined in Assumption 1(i). Then, we have

$$\sup_{0\le z\le 1} \|Q_n(z)\| = O_P(n\sqrt{h\log n}). \tag{3.2}$$

This theorem shows that the uniform rate (3.2) is exactly the same as (2.8) and therefore exceeds the stationary rate $O_P(\sqrt{nh\log n})$. This rate is also common across coordinates unlike the different rates that apply in the fixed design model. The reason for this common uniform rate is that there is no particular direction of dominance (such as $q_{\gamma_n(z)}$ in the fixed design case) because the random components Z_t are independently distributed with continuous density. The result is used in Section 4 to derive a uniform convergence rate for nonparametric kernel-based estimation of the functional coefficients in nonlinear cointegration models.

4 Cointegration models with varying coefficients

In this section we use the results developed earlier to derive the uniform consistency rate results for nonparametric kernel estimators in a nonlinear cointegration model with varying coefficients. The model has the form

$$Y_t = X_t'\beta(Z_t) + e_t, \quad t = 1, \cdots, n, \tag{4.1}$$

where X_t and e_t satisfy Assumption 1, $\beta(\cdot)$ is a d-dimensional coefficient function, and Z_t is either a fixed design or random design variate. In the fixed design case, model (4.1) is a cointegration model with time-varying coefficients, which has been studied in Park and Hahn (1999) and Phillips *et al* (2013). The model can then be regarded as an extension of the locally stationary models used in Robinson (1989) and Cai (2007) where the regressors are stationary. In the random design case, model (4.1) is a cointegration model with functional coefficients of the type studied in Cai *et al* (2009), Xiao (2009) and Gao and Phillips (2013b). These studies provide nonstationary extensions of the models considered in Fan and Zhang (1999) and Cai *et al* (2000). The existing literature in these cases focuses on the development of pointwise asymptotic theory for nonparametric estimators of the coefficient function $\beta(\cdot)$ (c.f., Cai *et al*, 2009; Phillips *et al*, 2013). Uniform consistency results and associated convergence rates in the nonstationary case have so far not been considered due to the technical difficulties involved in the presence of nonstationary regressors. This section aims to fill this gap in the literature.

Under a smoothness condition on $\beta(\cdot)$ and for some fixed z, we have the local approximation $\beta\left(Z_t\right) \approx \beta(z)$ when Z_t is in a small neighborhood of z. The kernel-weighted local level regression estimator of the coefficient $\beta(z)$ at z has the following form

$$\widehat{\beta}_n(z) = \Big[\sum_{t=1}^{n} X_t X_t' K\big(\frac{Z_t - z}{h}\big)\Big]^{+}\Big[\sum_{t=1}^{n} X_t Y_t K\big(\frac{Z_t - z}{h}\big)\Big], \tag{4.2}$$

where $\mathbf{A}^{+}$ denotes the Moore-Penrose inverse of $\mathbf{A}$, and as in the previous sections, $K(\cdot)$ is a kernel function and h is a bandwidth. We provide below a uniform consistency rate result for the estimator $\widehat{\beta}_n(z)$ over a range of values of z. Other kernel-based approaches such as local polynomial regression are also applicable to estimate the coefficient functions, and similar uniform consistency results as those given here can be obtained with some modification of the proofs.

To establish the limit theory for $\widehat{\beta}_n(\cdot)$, we impose the following commonly used smoothness condition on $\beta(\cdot)$ (c.f., Wang and Phillips, 2009a; Phillips *et al*, 2013).

Assumption 4. *The coefficient function $\beta(\cdot)$ is continuous with $\|\beta(z_1) - \beta(z_2)\| = O(|z_1 - z_2|^{\alpha_0})$ for $1/2 < \alpha_0 \leq 1$ and any $z_1, z_2 \in (0,1)$.*

We start with the fixed design case where $Z_t = \frac{t}{n}$ for $t = 1, \cdots, n$. Let $B_{z,*}(\mathbf{\Omega}_\varepsilon)$ be an independent copy of the d-dimensional Brownian motion $B_z(\mathbf{\Omega}_\varepsilon)$ which is defined as in (2.2). Define $\overline{b}_z = B_z(\mathbf{\Omega}_\varepsilon)$ and $\overline{q}_z = \overline{b}_z/\|\overline{b}_z\|$, and let $\overline{q}_z^{\perp}$ be the $d \times (d-1)$ orthogonal complement matrix such that

$$\mathbf{D}(z)'\mathbf{D}(z) = \mathbf{I}_d \;\text{ with }\; \mathbf{D}(z) = \big[\overline{q}_z, \overline{q}_z^{\perp}\big],$$

which can be seen as the limiting version of (2.4). Define

$$\mathbf{\Delta}_z = \begin{bmatrix} \Delta_1(z) & \Delta_2(z) \\ \Delta_2(z)' & \Delta_3(z) \end{bmatrix}, \tag{4.3}$$

with $\Delta_1(z) = \bar{b}_z'\bar{b}_z$,

$$\Delta_2(z) = \sqrt{2}\left(\bar{b}_z'\bar{b}_z\right)^{1/2}\left\{\int_{-1}^{1} B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)'K(r)dr\right\}\bar{q}_z^\perp,$$

and

$$\Delta_3(z) = 2(\bar{q}_z^\perp)'\left\{\int_{-1}^{1} B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)'K(r)dr\right\}\bar{q}_z^\perp.$$

Letting $\mathbf{R}_n$ and $\mathbf{D}_n(z)$ be defined as in Section 2, Proposition A.1 in Phillips *et al* (2013) shows that the standardized denominator matrix of (4.2) converges weakly to the limit:

$$\mathbf{R}_n^{-1}\mathbf{D}_n(z)'\Big[\sum_{t=1}^{n} X_tX_t'K\big(\frac{t-nz}{nh}\big)\Big]\mathbf{D}_n(z)\mathbf{R}_n^{-1} \Rightarrow \mathbf{\Delta}_z,$$

on which we make the following assumption.

Assumption 5. *The limit matrix $\mathbf{\Delta}_z$ is non-singular with probability 1 for any $z \in [\epsilon_\diamond, 1-\epsilon_\diamond]$, where $0 < \epsilon_\diamond < 1/2$ can be arbitrarily small.*

We provide the following discussion to justify this assumption. Using the technique of the triangular representation (Phillips, 1991), we may left transform $\mathbf{\Delta}_z$ by the nonsingular matrix

$$\begin{bmatrix} 1 & \mathbf{0}' \\ -W(z) & \mathbf{I}_{d-1} \end{bmatrix},$$

where $W(z) = \Delta_2(z)'/\Delta_1(z) = \sqrt{2}\left(\bar{b}_z'\bar{b}_z\right)^{-1/2}(\bar{q}_z^\perp)'\left\{\int_{-1}^{1} B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)K(r)dr\right\}$, giving the following matrix

$$\mathbf{\Delta}_z^* = \begin{bmatrix} 1 & \mathbf{0}' \\ -W(z) & \mathbf{I}_{d-1} \end{bmatrix}\begin{bmatrix} \Delta_1(z) & \Delta_2(z) \\ \Delta_2(z)' & \Delta_3(z) \end{bmatrix} = \begin{bmatrix} \Delta_1(z) & \Delta_2(z) \\ \mathbf{0} & \Delta_3^*(z) \end{bmatrix}$$

where

$$\Delta_3^*(z) = 2(\bar{q}_z^\perp)'\left\{\int B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)'K(r)dr - \Big[\int B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)K(r)dr\Big]\Big[\int B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)'K(r)dr\Big]\right\}\bar{q}_z^\perp,$$

in which we use the notation $\int \equiv \int_{-1}^{1}$. Note that $\Delta_1(z)$ is positive with probability 1 if $z \geq \epsilon_\diamond$. Hence, if $\int B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)'K(r)dr - \Big[\int B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)K(r)dr\Big]\Big[\int B_{\frac{r+1}{2},*}(\mathbf{\Omega}_\varepsilon)'K(r)dr\Big]$ is non-singular with probability 1, $\mathbf{\Delta}_z^*$ is non-singular for any $z \in [\epsilon_\diamond, 1-\epsilon_\diamond]$, which justifies Assumption 5.

Based on Theorem 2.1 and Corollary 2.1, we obtain the following uniform consistency rate results for the kernel estimator $\widehat{\beta}_n(z)$.

THEOREM 4.1. *Suppose that the assumptions in Theorem 2.1 and Assumptions 4 and 5 are satisfied. Then, we have as $n \to \infty$*

$$\sup_{\epsilon_\diamond \le z \le 1-\epsilon_\diamond} \left| q'_{\gamma_n(z)} \left[\widehat{\beta}_n(z) - \beta(z) \right] \right| = O_P\Big(h^{\alpha_0} + \sqrt{\frac{\log n}{n^2 h}} \Big) \tag{4.4}$$

and

$$\sup_{\epsilon_\diamond \le z \le 1-\epsilon_\diamond} \left\| (q^\perp_{\gamma_n(z)})' \left[\widehat{\beta}_n(z) - \beta(z) \right] \right\| = O_P\Big(h^{\alpha_0} + \frac{\sqrt{\log n}}{nh} \Big), \tag{4.5}$$

where $\epsilon_\diamond$ is defined in Assumption 5.

Note that

$$\begin{aligned}
\widehat{\beta}_n(z) - \beta(z) \quad = \quad & \Big[\sum_{t=1}^n X_t X_t' K\big(\frac{Z_t - z}{h}\big) \Big]^+ \Big[\sum_{t=1}^n X_t X_t' [\beta(Z_t) - \beta(z)] K\big(\frac{Z_t - z}{h}\big) \Big] + \\
& \Big[\sum_{t=1}^n X_t X_t' K\big(\frac{Z_t - z}{h}\big) \Big]^+ \Big[\sum_{t=1}^n X_t e_t K\big(\frac{Z_t - z}{h}\big) \Big].
\end{aligned} \tag{4.6}$$

The order $O_P(h^{\alpha_0})$ is contributed by the first term on the right hand side of (4.6), which measures the bias effect of the nonparametric estimator $\widehat{\beta}_n(z)$. This order can be improved to $O_P(h^2)$ if the local linear method (c.f., Fan and Gijbels, 1996) is used to estimate $\beta(\cdot)$. We next discuss the twin uniform convergence rates contributed by the second term on the right hand side of (4.6). Theorem 4.1 above gives different uniform convergence rates in the two directions determined by the kernel degeneracy, just as in Corollary 2.1. In the direction $q_{\gamma_n(z)}$, we have the uniform convergence rate $O_P\big(\sqrt{\frac{\log n}{n^2 h}}\big)$, which we call the *type I uniform convergence rate*. This rate is faster than the rate $O_P\left(\frac{\sqrt{\log n}}{nh}\right)$ that applies in the other direction (c.f. (4.5)) as well as the usual rate $O_P\big(\sqrt{\frac{\log n}{nh}}\big)$ that applies in the stationary case. In the direction $q^\perp_{\gamma_n(z)}$, the uniform convergence rate $O_P\big(\frac{\sqrt{\log n}}{nh}\big)$ is slower than the type I uniform convergence rate of (4.4), but is still faster than the stationary rate. The rate $O_P\big(\frac{\sqrt{\log n}}{nh}\big)$ is called the *type II uniform convergence rate*.

Next consider the random design case where the covariate Z_t is *i.i.d.*, as discussed in Section 3. Define

$$\mathbf{\Lambda}_z = f_Z(z) \int_0^1 B_r(\mathbf{\Omega}_\varepsilon) B_r(\mathbf{\Omega}_\varepsilon)' dr,$$

where $f_Z(\cdot)$ is the density function of Z_t. Similar to the argument in the proof of Proposition A.1 in Li *et al* (2014), it is easy to show that

$$\frac{1}{n^2 h}\sum_{t=1}^{n} X_t X_t' K\left(\frac{Z_t - z}{h}\right) \Rightarrow \mathbf{\Lambda}_z$$

for $0 \leq z \leq 1$, and $\mathbf{\Lambda}_z$ is non-singular if $f_Z(\cdot)$ is strictly positive by using Lemma A2 in Phillips and Hansen (1990). Hence, there is no kernel degeneracy issue for this case. Using Theorem 3.1 we derive a uniform convergence rate for $\widehat{\beta}_n(\cdot)$ in the following theorem, which shows that a common type I uniform convergence rate is attained in all directions in the random design case.

THEOREM 4.2. *Suppose that the assumptions in Theorem 3.1 and Assumption 4 are satisfied. Let the density function $f_Z(z)$ be bounded away from zero and infinity for all $z \in [0,1]$. Then, we have as $n \to \infty$*

$$\sup_{0\leq z\leq 1} \|\widehat{\beta}_n(z) - \beta(z)\| = O_P\Big(h^{\alpha_0} + \sqrt{\frac{\log n}{n^2 h}}\Big). \tag{4.7}$$

This uniform consistency result gives a sharp rate of convergence for estimation of nonlinear cointegration models with functional coefficients and complements the pointwise limit theory developed by Cai *et al* (2009), Xiao (2009) and Gao and Phillips (2013b).

5 Conclusions

This paper has derived uniform consistency results for nonparametric kernel-weighted sample covariances and regressions in a nonstationary data framework. This framework has practical applications in varying coefficient regressions with coefficient covariates that follow either fixed or random designs. In the fixed design case, two different uniform asymptotic rates have been obtained, depending on a certain covariate-sensitive random direction, a result that is quite different from the random design case where a common uniform asymptotic rate applies. Both results are shown to be robust to endogeneity of the regressors.

A regression application of these results has confirmed the uniform super-consistency of nonparametric kernel estimates of the coefficient functions in nonlinear cointegration models with varying coefficients and gives sharp convergence rates in this regression case. In the fixed design framework, two types of uniform convergence rates again have been established in the covariate sensitive random directions and both rates are faster than the rate in the stationary case. In the random design framework, there is a common uniform convergence rate, which is also faster than that of the stationary case. These uniform consistency results are relevant in estimating semiparametric cointegration models with partially-varying coefficients, long run variance estimation in such models, kernel-based specification testing of nonlinear cointegration models, and the theory for the optimal bandwidth selection in the nonparametric kernel-smoothing under nonstationarity.

6 Acknowledgements

The authors thank the Co-Editor, Professor Oliver Linton, and three referees for their helpful and constructive comments. Thanks also go to Dr Jia Chen for useful suggestions. Phillips acknowledges support from the NSF under Grant No. SES 12-58258. Gao acknowledges support from the Australian Research Council Discovery Grants Program under Grant Nos. DP1096374 and DP130104229.

References

Billingsley, P. (1968). *Convergence of Probability Measure*. Wiley, New York.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction* (2nd ed.). Lecture Notes in Statistics 110. Springer–Verlag.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* **136**, 163–188.

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying–coefficient models. *Journal of American Statistical Association* **95**, 888–902.

Cai, Z., Li, Q. and Park, J. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* **148**, 101–113.

Chan, N. and Wang, Q. (2012). Uniform convergence for Nadaraya-Watson estimators with non-stationary data. Working paper, School of Mathematics and Statistics, University of Sydney, available at http://www.maths.usyd.edu.au/u/pubs/publist/preprints/2012/chan-12.pdf

Chen, J., Gao, J. and Li, D. (2012). Estimation in semiparametric regression with nonstationary regressors. *Bernoulli* **18**, 678–702.

Chen, J., Li, D. and Zhang, L. (2010). Robust estimation in a nonlinear cointegration model. *Journal of Multivariate Analysis* **101**, 706–717.

Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press.

Duffy, J. (2013). Uniform convergence rates, on a maximal domain, for structural nonparametric cointegrating regression. Unpublished paper, Yale University.

Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics* **27**, 1491–1518.

Gao, J., Kanaya, S., Li, D. and Tjøstheim, D. (2015). Uniform consistency for nonparametric estimators in null recurrent time series. Forthcoming in *Econometric Theory*.

Gao, J. and Phillips, C. B. P. (2013a). Semiparametric estimation in triangular system equations with nonstationarity. *Journal of Econometrics* **176**, 59–79.

Gao, J. and Phillips, C. B. P. (2013b). Functional coefficient nonstationary regression. *Cowles Foundation Discussion Paper* **1911**. Cowles Foundation for Research in Economics, Yale University, available at http://cowles.econ.yale.edu/P/cd/d19a/d1911.pdf.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.

Ibragimov, R. and Phillips, P. C. B. (2008). Regression asymptotics using martingale convergence methods. *Econometric Theory* **24**, 888–947.

Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics* **29**, 372–416.

Karlsen, H. A., Myklebust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* **35**, 252–299.

Kristensen, D. (2009). Uniform convergence rates of kernel estimators with heterogenous dependent data. *Econometric Theory* **25**, 1433–1445.

Li, D., Lu, Z. and Linton, O. (2012). Local linear fitting under near epoch dependence: uniform consistency with convergence rate. *Econometric Theory* **28**, 935–958.

Li, K., Li, D., Liang, Z. and Hsiao, C. (2014). Estimation of semi-varying coefficient models with nonstationary regressors. Forthcoming in *Econometric Reviews*.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications* **65**, 69–80.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichskeittheorie und verwandte Gebiete* **61**, 405-415.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis* **17**, 571–599.

Park, J. Y. and Hahn, S. B. (1999). Cointegrating regressions with time varying coefficients. *Econometric Theory* **15**, 664–703.

de la Peña, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *Annals of Probability* **27**, 537–564.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica* **59**, 283-306.

Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies* **53**, 473-496.

Phillips, P. C. B. and Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* **57**, 99-125.

Phillips, P. C. B., Li, D. and Gao, J. (2013). Estimating smooth structural change in cointegration models. *Cowles Foundation Discussion Paper* **1910**, Cowles Foundation for Research in Economics, Yale University, available at http://cowles.econ.yale.edu/P/cd/d19a/d1910.pdf.

Phillips, P. C. B. and Park J. (1998). Nonstationary density estimation and kernel autoregression. *Cowles Foundation Discussion Paper* **1181**, Cowles Foundation for Research in Economics, Yale University, available at http://cowles.econ.yale.edu/P/cd/d11b/d1181.pdf.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics* **20**, 971–1001.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. *Statistical Analysis and Forecasting of Economic Structural Change* (ed. by P. Hackl). Springer, Berlin, pp. 164–253.

Roussas, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Processes and Their Applications* **36**, 107–116.

van der Vaart, A. W. and Wellner, J. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer.

Wang, Q. and Chan, N. (2014). Uniform convergence rates for a class of martingales with application in non-linear cointegrating regression. *Bernoulli* **20**, 207-230.

Wang, Q. and Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25**, 710-738.

Wang, Q. and Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica* **77**, 1901-1948.

Wang, Q. and Wang, Y. (2013). Nonparametric cointegrating regression with NNH errors. *Econometric Theory* **29**, 1–27.

Xiao, Z. (2009). Functional-coefficient cointegrating regression. *Journal of Econometrics* **152**, 81-92.

A Proofs of the main results

This appendix provides proofs of the main results in Sections 2–4. To simplify notation, in the sequel we let $q_z = q_{\gamma_n(z)}$, $q_z^{\perp} = q_{\gamma_n(z)}^{\perp}$, and use C for a positive constant whose value may change according to its position.

Proof of Theorem 2.1. For $\epsilon_* \le z \le 1-\epsilon_*$, define

$$\begin{aligned} Q_n(z,1) &= \frac{q_z'}{n\sqrt{h}}\sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)X_t e_t, \\ Q_n(z,2) &= \frac{(q_z^{\perp})'}{nh}\sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)X_t e_t. \end{aligned}$$

Note that

$$Q_n(z,1) = \frac{q_z'}{n\sqrt{h}}X_{\gamma_n(z)}\sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)e_t + \frac{q_z'}{n\sqrt{h}}\sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)(X_t - X_{\gamma_n(z)})e_t, \tag{A.1}$$

where $\gamma_n(z)$ is defined in Section 2, and

$$Q_n(z,2) = \frac{(q_z^{\perp})'}{nh}\sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)(X_t - X_{\gamma_n(z)})e_t, \tag{A.2}$$

as $q_z^{\perp}$ is orthogonal to $X_{\gamma_n(z)}$ by (2.4) in Section 2. By continuous mapping (e.g. Billingsley, 1968), it is easy to show that

$$\sup_{\epsilon_* \le z \le 1-\epsilon_*}\big(\|q_z\| + \|q_z^{\perp}\|\big) = O_P(1). \tag{A.3}$$

Then, by (A.1)–(A.3), it is sufficient to show that

$$\sup_{\epsilon_* \le z \le 1-\epsilon_*}\left|\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)e_t\right| = O_P(\sqrt{\log n}), \tag{A.4}$$

and

$$\sup_{\epsilon_* \le z \le 1-\epsilon_*}\left\|\frac{1}{nh}\sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)(X_t - X_{\gamma_n(z)})e_t\right\| = O_P(\sqrt{\log n}), \tag{A.5}$$

which we now prove in turn.

Proof of (A.4). Using the so-called Beveridge-Nelson (BN) decomposition approach (c.f., Phillips and Solo, 1992):

$$\phi(\mathcal{L}) \equiv \sum_{j=0}^{\infty}\phi_j\mathcal{L}^j = \sum_{j=0}^{\infty}\phi_j - (1-\mathcal{L})\sum_{j=0}^{\infty}\widetilde{\phi}_j\mathcal{L}^j \equiv \phi - (1-\mathcal{L})\widetilde{\phi}(\mathcal{L})$$

with $\widetilde{\phi}_j = \sum_{k=j+1}^{\infty}\phi_k$, we have

$$e_t = \overline{e}_t + (\widetilde{e}_{t-1} - \widetilde{e}_t), \tag{A.6}$$

where $\overline{e}_t = \big(\sum_{j=0}^{\infty}\phi_j\big)\eta_t = \phi\eta_t$ and $\widetilde{e}_t = \sum_{j=0}^{\infty}\widetilde{\phi}_j\eta_{t-j}$. By (A.6), we have

$$\begin{aligned}
\sum_{t=1}^{n} e_t K\big(\frac{t-nz}{nh}\big) &= \sum_{t=1}^{n}\overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n}\widetilde{e}_{t-1} K\big(\frac{t-nz}{nh}\big) - \sum_{t=1}^{n}\widetilde{e}_t K\big(\frac{t-nz}{nh}\big)\\
&= \sum_{t=1}^{n}\overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n}\widetilde{e}_{t-1} K\big(\frac{t-1-nz}{nh}\big) - \sum_{t=1}^{n}\widetilde{e}_t K\big(\frac{t-nz}{nh}\big) +\\
&\quad \sum_{t=1}^{n}\widetilde{e}_{t-1}\left[K\big(\frac{t-nz}{nh}\big) - K\big(\frac{t-1-nz}{nh}\big)\right]\\
&= \sum_{t=1}^{n}\overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n}\widetilde{e}_{t-1}\left[K\big(\frac{t-nz}{nh}\big) - K\big(\frac{t-1-nz}{nh}\big)\right] +\\
&\quad \widetilde{e}_0 K\big(\frac{-z}{h}\big) - \widetilde{e}_n K\big(\frac{1-z}{h}\big).
\end{aligned}$$

By virtue of Assumption 2(i) and (ii),

$$\widetilde{e}_0 K\big(\frac{-z}{h}\big) = \widetilde{e}_n K\big(\frac{1-z}{h}\big) = 0 \tag{A.7}$$

with probability 1 for any $\epsilon_* \le z \le 1-\epsilon_*$, which indicates that

$$\sum_{t=1}^{n} e_t K\big(\frac{t-nz}{nh}\big) = \sum_{t=1}^{n}\overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n}\widetilde{e}_{t-1}\left[K\big(\frac{t-nz}{nh}\big) - K\big(\frac{t-1-nz}{nh}\big)\right] \tag{A.8}$$

uniformly for $\epsilon_* \le z \le 1-\epsilon_*$.

Define $\mathbb{S}_k = \{s\big|(k-1)nr_n + 1 \le s < knr_n\}$ for $k = 1, 2, \cdots, R_n$, and $\mathbb{S}_{R_n+1} = \{s\big|nr_nR_n + 1 \le s \le n\}$, where $R_n = \lfloor r_n^{-1}\rfloor$, $r_n = n^{-1/2}h^{3/2}\log^{1/2}(n)$. Let s_k be the smallest number in the set $\mathbb{S}_k$ for $k = 1, \cdots, R_n, R_n+1$, and $R_n^* = R_n + 1$. By standard arguments, we have for n large enough,

$$\begin{aligned}
\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left|\sum_{t=1}^{n}\overline{e}_t K\big(\frac{t-nz}{nh}\big)\right| &\le \max_{1\le k\le R_n^*}\sup_{s\in\mathbb{S}_k}\left|\sum_{t=1}^{n}\overline{e}_t\big[K\big(\frac{t-s}{nh}\big) - K\big(\frac{t-s_k}{nh}\big)\big]\right| +\\
&\quad \max_{1\le k\le R_n^*}\left|\sum_{t=1}^{n}\overline{e}_t K\big(\frac{t-s_k}{nh}\big)\right|.
\end{aligned}$$

Noting that

$$\sup_{s\in\mathbb{S}_k}\left|K\big(\frac{t-s}{nh}\big) - K\big(\frac{t-s_k}{nh}\big)\right| \le C\sup_{s\in\mathbb{S}_k}\left|\frac{s-s_k}{nh}\right| \le C\cdot\frac{nr_n}{nh} = Cr_nh^{-1}$$

and $|\overline{e}_t| = O_P(1)$, we deduce that

$$\max_{1\le k\le R_n^*}\sup_{s\in\mathbb{S}_k}\left|\sum_{t=1}^{n}\overline{e}_t\left[K\big(\frac{t-s}{nh}\big) - K\big(\frac{t-s_k}{nh}\big)\right]\right| = n\cdot O_P(r_nh^{-1}) = O_P\big(\frac{nr_n}{h}\big) = O_P\big(\sqrt{nh\log n}\big). \tag{A.9}$$

Noting that $\bar{e}_t = \phi\eta_t$, we next prove

$$\max_{1\le k\le R_n^*}\left|\sum_{t=1}^{n}\eta_t K\big(\frac{t-s_k}{nh}\big)\right| = O_P\big(\sqrt{nh\log n}\big) \tag{A.10}$$

by the truncation technique and using the Bernstein inequality. Let

$$\overline{\eta}_t = \eta_t\cdot \mathrm{I}\Big(|\eta_t|\le \frac{(nh)^{1/2}}{(\log n)^{1/2}}\Big) \quad\text{and}\quad \widetilde{\eta}_t = \eta_t-\overline{\eta}_t = \eta_t\cdot \mathrm{I}\Big(|\eta_t| > \frac{(nh)^{1/2}}{(\log n)^{1/2}}\Big),$$

where $\mathrm{I}(\cdot)$ is an indicator function. Note that

$$\mathbb{E}\left[|\widetilde{\eta}_t|\right] = \mathbb{E}\left[|\eta_t|\mathrm{I}\Big(|\eta_t| > \frac{(nh)^{1/2}}{(\log n)^{1/2}}\Big)\right] \le \frac{(\log n)^{(3+\delta_0)/2}}{(nh)^{(3+\delta_0)/2}}\cdot\mathbb{E}\left[|\eta_t|^{4+\delta_0}\right],$$

which indicates that

$$\max_{1\le k\le R_n^*}\left|\sum_{t=1}^{n}\mathbb{E}[\widetilde{\eta}_t]K\big(\frac{t-s_k}{nh}\big)\right| = O_P\left(\frac{(\log n)^{(3+\delta_0)/2}}{(nh)^{(1+\delta_0)/2}}\right) = o_P\big(\sqrt{nh\log n}\big).$$

Hence, in order to prove

$$\max_{1\le k\le R_n^*}\left|\sum_{t=1}^{n}\big(\widetilde{\eta}_t-\mathbb{E}[\widetilde{\eta}_t]\big)K\big(\frac{t-s_k}{nh}\big)\right| = o_P\big(\sqrt{nh\log n}\big), \tag{A.11}$$

we need only to show that

$$\max_{1\le k\le R_n^*}\left|\sum_{t=1}^{n}\widetilde{\eta}_t K\big(\frac{t-s_k}{nh}\big)\right| = o_P\big(\sqrt{nh\log n}\big). \tag{A.12}$$

Notice that if $|\eta_t|\le \frac{(nh)^{1/2}}{(\log n)^{1/2}}$ holds for all $t=1,\cdots,n$, we have $\widetilde{\eta}_t\equiv 0$ and thus $\big|\sum_{t=1}^{n}\widetilde{\eta}_t K\big(\frac{t-s_k}{nh}\big)\big| = 0$ for any $k=1,\cdots,R_n^*$, which indicates that

$$\left\{\max_{1\le k\le R_n^*}\left|\sum_{t=1}^{n}\widetilde{\eta}_t K\big(\frac{t-s_k}{nh}\big)\right|\neq 0\right\}\subset\left\{\max_{1\le t\le n}|\eta_t| > \frac{(nh)^{1/2}}{(\log n)^{1/2}}\right\}.$$

Using the above fact and noting that

$$\mathbb{P}\left\{\max_{1\le t\le n}|\eta_t| > \frac{(nh)^{1/2}}{(\log n)^{1/2}}\right\}\le C\cdot\frac{n(\log n)^{(4+\delta_0)/2}}{(nh)^{(4+\delta_0)/2}} = o(1)$$

as $\frac{n^{2+\delta_0}h^{4+\delta_0}}{(\log n)^{4+\delta_0}}\to\infty$, we can complete the proof of (A.12).

On the other hand, note that $\{\eta_t\}$ is a sequence of *i.i.d.* random variables, and the number of non-zero summands in $\sum_{t=1}^{n}\overline{\eta}_t K\big(\frac{t-s_k}{nh}\big)$ is of order (nh) as the compact support of the kernel function is $[-1,1]$. Let c_K be a positive constant which is the upper bound for the kernel function $K(\cdot)$. Letting c_0 be some positive constant such that

$$\max_{1\le k\le R_n^*} \mathrm{Var}\left(\sum_{t=1}^{n} \overline{\eta}_t K\big(\frac{t-s_k}{nh}\big)\right) = \max_{1\le k\le R_n^*} \mathrm{Var}\left(\sum_{t=s_k-nh}^{s_k+nh} \overline{\eta}_t K\big(\frac{t-s_k}{nh}\big)\right) \le c_0 nh,$$

and by using a Bernstein type inequality (e.g., Lemma 2.2.9 in van der Vaart and Wellner, 1996), we have

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\le k\le R_n^*}\left|\sum_{t=1}^{n}\big(\overline{\eta}_t-\mathbb{E}[\overline{\eta}_t]\big)K\big(\frac{t-s_k}{nh}\big)\right| > M\sqrt{nh\log n}\right\}\\
\le\ & \sum_{k=1}^{R_n^*}\mathbb{P}\left\{\left|\sum_{t=1}^{n}\big(\overline{\eta}_t-\mathbb{E}[\overline{\eta}_t]\big)K\big(\frac{t-s_k}{nh}\big)\right| > M\sqrt{nh\log n}\right\}\\
\le\ & \sum_{k=1}^{R_n^*} 2\exp\left\{-\frac{M^2 nh\log n}{(2c_0+4c_K M/3)nh}\right\} = O\left(r_n^{-1} n^{-\sqrt{M}}\right) = o(1),
\end{aligned}$$

where M is chosen such that

$$M^{3/2} > \frac{4c_K M}{3} + 2c_0, \quad r_n^{-1} n^{-\sqrt{M}} = o(1),$$

which are possible when M is sufficiently large. The above calculation indicates that

$$\max_{1\le k\le R_n^*}\left|\sum_{t=1}^{n}\big(\overline{\eta}_t-\mathbb{E}[\overline{\eta}_t]\big)K\big(\frac{t-s_k}{nh}\big)\right| = O_P\left(\sqrt{nh\log n}\right). \tag{A.13}$$

Then, by (A.11) and (A.13), we can prove (A.10), which together with (A.9), leads to

$$\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left|\sum_{t=1}^{n}\overline{e}_t K\big(\frac{t-nz}{nh}\big)\right| = O_P\left(\sqrt{nh\log n}\right). \tag{A.14}$$

Noting that $\left|K\big(\frac{t-nz}{nh}\big) - K\big(\frac{t-1-nz}{nh}\big)\right| \le C\frac{1}{nh}$ and $|\widetilde{e}_t| = o_P((nh)^{1/4})$ uniformly in $t=1,\cdots,n$, by a standard derivation, we can also show that

$$\begin{aligned}
&\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left|\sum_{t=1}^{n}\widetilde{e}_{t-1}\left[K\big(\frac{t-nz}{nh}\big)-K\big(\frac{t-1-nz}{nh}\big)\right]\right|\\
=\ & \sup_{\epsilon_*\le z\le 1-\epsilon_*}\left|\sum_{t=\lfloor n(z-h)\rfloor}^{\lfloor n(z+h)\rfloor+2}\widetilde{e}_{t-1}\left[K\big(\frac{t-nz}{nh}\big)-K\big(\frac{t-1-nz}{nh}\big)\right]\right|\\
\le\ & (2\lfloor nh\rfloor+2)\cdot C(nh)^{-1}\cdot\max_{1\le t\le n}|\widetilde{e}_t|\\
=\ & o_P\left(\sqrt{nh\log n}\right),
\end{aligned} \tag{A.15}$$

which together with (A.7), (A.8) and (A.14), leads to (A.4).

Proof of (A.5). Using the BN decomposition again, we have

$$X_t - X_{\gamma_n(z)} = \sum_{s=\gamma_n(z)+1}^{t} v_s = \sum_{s=\gamma_n(z)+1}^{t} \overline{v}_s + \widetilde{v}_{\gamma_n(z)} - \widetilde{v}_t,$$

where $\overline{v}_t = \left(\sum_{j=0}^{\infty} \mathbf{\Phi}_j\right) \varepsilon_t = \mathbf{\Phi}\varepsilon_t$ and $\widetilde{v}_t = \sum_{j=0}^{\infty} \widetilde{\mathbf{\Phi}}_j \varepsilon_{t-j}$ with $\widetilde{\mathbf{\Phi}}_j = \sum_{k=j+1}^{\infty} \mathbf{\Phi}_k$. Hence, in order to prove (A.5), we need only prove that

$$\sum_{t=1}^{n} \Big(\sum_{s=\gamma_n(z)+1}^{t} \overline{v}_s \Big) e_t K\big(\frac{t-nz}{nh}\big) = O_P\left(nh\sqrt{\log n}\right), \tag{A.16}$$

$$\widetilde{v}_{\gamma_n(z)} \sum_{t=1}^{n} e_t K\big(\frac{t-nz}{nh}\big) = o_P\left(nh\sqrt{\log n}\right), \tag{A.17}$$

$$\sum_{t=1}^{n} \widetilde{v}_t e_t K\big(\frac{t-nz}{nh}\big) = o_P\left(nh\sqrt{\log n}\right), \tag{A.18}$$

uniformly for $\epsilon_* \le z \le 1-\epsilon_*$.

Note that both $\widetilde{v}_t$ and e_t are well defined stationary linear processes, and the numbers of non-zero summands in both $\sum_{t=1}^{n} \widetilde{v}_t e_t K\big(\frac{t-nz}{nh}\big)$ and $\sum_{t=1}^{n} e_t K\big(\frac{t-nz}{nh}\big)$ are of order (nh). We can thus prove (A.17) and (A.18) by arguments similar to those in the proof of (A.4) above. This leaves (A.16), which will be proved next.

In order to prove (A.16) we proceed as follows. Let $\overline{v}_t(z) = \sum_{s=\gamma_n(z)+1}^{t} \overline{v}_s$ and $\overline{v}_t(z) = 0$ if $t < \gamma_n(z)+1$. Using the BN decomposition (A.6) and some basic algebra, we have

$$\begin{aligned}
\sum_{t=1}^{n} \overline{v}_t(z) e_t K\big(\frac{t-nz}{nh}\big) &= \sum_{t=1}^{n} \overline{v}_t(z) \overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n} \overline{v}_t(z) \widetilde{e}_{t-1} K\big(\frac{t-nz}{nh}\big) - \\
&\quad \sum_{t=1}^{n} \overline{v}_t(z) \widetilde{e}_t K\big(\frac{t-nz}{nh}\big) \\
&= \sum_{t=1}^{n} \overline{v}_t \overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n} \overline{v}_{t-1}(z) \overline{e}_t K\big(\frac{t-nz}{nh}\big) + \\
&\quad \sum_{t=1}^{n} \overline{v}_{t-1}(z) \widetilde{e}_{t-1} \left[K\big(\frac{t-nz}{nh}\big) - K\big(\frac{t-1-nz}{nh}\big) \right] - \\
&\quad \overline{v}_n(z) \widetilde{e}_n K\big(\frac{1-z}{h}\big) + \sum_{t=1}^{n} \overline{v}_t \widetilde{e}_{t-1} K\big(\frac{t-nz}{nh}\big).
\end{aligned}$$

Similar to the proof of (A.15), noting that $\left|K\big(\frac{t-nz}{nh}\big) - K\big(\frac{t-1-nz}{nh}\big)\right| \le C\frac{1}{nh}$, and $|\widetilde{e}_t| = o_P\left((nh)^{1/4}\right)$ and $|\overline{v}_t(z)| = o_P\left((nh)^{3/4}\right)$ uniformly in $\epsilon_* \le z \le 1-\epsilon_*$ and $n(z-h) < t < n(z+h)$, by a standard derivation we may show that

$$\begin{aligned}
&\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\overline{v}_{t-1}(z)\widetilde{e}_{t-1}\left[K\big(\frac{t-nz}{nh}\big)-K\big(\frac{t-1-nz}{nh}\big)\right]\right\| \\
= &\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=\lfloor n(z-h)\rfloor}^{\lfloor n(z+h)\rfloor+2}\overline{v}_{t-1}(z)\widetilde{e}_{t-1}\left[K\big(\frac{t-nz}{nh}\big)-K\big(\frac{t-1-nz}{nh}\big)\right]\right\| \\
\le &\ (2\lfloor nh\rfloor+2)\cdot C(nh)^{-1}\cdot\max_{1\le t\le n}\left(|\widetilde{e}_t|\sup_{\epsilon_*\le z\le 1-\epsilon_*}|\overline{v}_{t-1}(z)|\right) \\
= &\ o_P\left(nh\sqrt{\log n}\right). \qquad \text{(A.19)}
\end{aligned}$$

Note that

$$\begin{aligned}
\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\overline{v}_t\overline{e}_tK\big(\frac{t-nz}{nh}\big)\right\| \le &\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\left\{\overline{v}_t\overline{e}_t-\mathbb{E}[\overline{v}_t\overline{e}_t]\right\}K\big(\frac{t-nz}{nh}\big)\right\|+ \\
&\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\mathbb{E}[\overline{v}_t\overline{e}_t]K\big(\frac{t-nz}{nh}\big)\right\|.
\end{aligned}$$

By Assumptions 1 and 2, we then have

$$\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\mathbb{E}[\overline{v}_t\overline{e}_t]K\big(\frac{t-nz}{nh}\big)\right\| = O\left(nh\right)=o\left(nh\sqrt{\log n}\right). \qquad \text{(A.20)}$$

On the other hand, noting that $\{\overline{v}_t\overline{e}_t-\mathbb{E}[\overline{v}_t\overline{e}_t]\}$ is a sequence of *i.i.d.* random vectors with mean zero, and following the proof of (A.10), we can similarly prove that

$$\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\left\{\overline{v}_t\overline{e}_t-\mathbb{E}[\overline{v}_t\overline{e}_t]\right\}K\big(\frac{t-nz}{nh}\big)\right\| = O_P\left(\sqrt{nh\log n}\right)=o_P\left(nh\sqrt{\log n}\right). \qquad \text{(A.21)}$$

Then, a combination of (A.20) and (A.21) leads to

$$\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\overline{v}_t\overline{e}_tK\big(\frac{t-nz}{nh}\big)\right\| = o_P\left(nh\sqrt{\log n}\right). \qquad \text{(A.22)}$$

Noting that $K\big(\frac{1-z}{h}\big)\equiv 0$ for $\epsilon_*\le z\le 1-\epsilon_*$, we have that

$$\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\overline{v}_n(z)\widetilde{e}_nK\big(\frac{1-z}{h}\big)\right\| = o_P\left(nh\sqrt{\log n}\right). \qquad \text{(A.23)}$$

By Assumption 1(ii), $\overline{v}_t$ is independent of $\widetilde{e}_{t-1}$, which implies that $\left\{\overline{v}_t\widetilde{e}_{t-1}K\big(\frac{t-nz}{nh}\big):\ t=1,\cdots,n\right\}$ forms an array of martingale differences. Following the arguments in the proof of (A.10) with some modifications (e.g. replacing the *i.i.d.* Bernstein inequality by the exponential inequality for martingale differences in de la Peña, 1999), it follows that

$$\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\overline{v}_t\widetilde{e}_{t-1}K\big(\frac{t-nz}{nh}\big)\right\| = o_P\left(nh\sqrt{\log n}\right). \tag{A.24}$$

By (A.19) and (A.22)–(A.24), in order to complete the proof of (A.16), we need only prove that

$$\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\overline{v}_{t-1}(z)\overline{e}_tK\big(\frac{t-nz}{nh}\big)\right\| = O_P\left(nh\sqrt{\log n}\right). \tag{A.25}$$

Let $\mathbb{S}_k$ and s_k be defined as in the proof of (A.4) with r_n, R_n and R_n^* replaced by $\overline{r}_n = n^{-1}h^2\log n$, $\overline{R}_n = \lfloor \overline{r}_n^{-1}\rfloor$ and $\overline{R}_n^* = \overline{R}_n + 1$, respectively. By standard arguments, we have

$$\begin{aligned}
\sup_{\epsilon_*\le z\le 1-\epsilon_*}\left\|\sum_{t=1}^{n}\overline{v}_{t-1}(z)\overline{e}_tK\big(\frac{t-nz}{nh}\big)\right\| \le\ & \max_{1\le k\le \overline{R}_n^*}\sup_{s\in\mathbb{S}_k}\left\|\sum_{t=1}^{n}\overline{v}_{t-1,*}(s)\overline{e}_t\left[K\big(\frac{t-s}{nh}\big)-K\big(\frac{t-s_k}{nh}\big)\right]\right\| + \\
& \max_{1\le k\le \overline{R}_n^*}\sup_{s\in\mathbb{S}_k}\left\|\sum_{t=1}^{n}\left[\overline{v}_{t-1,*}(s)-\overline{v}_{t-1,*}(s_k)\right]\overline{e}_tK\big(\frac{t-s_k}{nh}\big)\right\| + \\
& \max_{1\le k\le \overline{R}_n^*}\left\|\sum_{t=1}^{n}\overline{v}_{t-1,*}(s_k)\overline{e}_tK\big(\frac{t-s_k}{nh}\big)\right\|, \\
\equiv\ & \Xi_{n1}+\Xi_{n2}+\Xi_{n3},
\end{aligned}$$

where $\overline{v}_{t,*}(s) = \overline{v}_t(s/n)$. In order to prove (A.25), we need to show that

$$\Xi_{n1} = \max_{1\le k\le \overline{R}_n^*}\sup_{s\in\mathbb{S}_k}\left\|\sum_{t=1}^{n}\overline{v}_{t-1,*}(s)\overline{e}_t\left[K\big(\frac{t-s}{nh}\big)-K\big(\frac{t-s_k}{nh}\big)\right]\right\| = O_P\left(nh\sqrt{\log n}\right), \tag{A.26}$$

$$\Xi_{n2} = \max_{1\le k\le \overline{R}_n^*}\sup_{s\in\mathbb{S}_k}\left\|\sum_{t=1}^{n}\left[\overline{v}_{t-1,*}(s)-\overline{v}_{t-1,*}(s_k)\right]\overline{e}_tK\big(\frac{t-s_k}{nh}\big)\right\| = O_P\left(nh\sqrt{\log n}\right) \tag{A.27}$$

and

$$\Xi_{n3} = \max_{1\le k\le \overline{R}_n^*}\left\|\sum_{t=1}^{n}\overline{v}_{t-1,*}(s_k)\overline{e}_tK\big(\frac{t-s_k}{nh}\big)\right\| = O_P\left(nh\sqrt{\log n}\right). \tag{A.28}$$

We next provide the proof of (A.28). By the definition of $\overline{e}_t$, we may prove (A.28) with $\overline{e}_t$ in Ξ_{n3} replaced by η_t. Let $w_t(s_k) = \overline{v}_{t-1,*}(s_k)\eta_t$, $\mathcal{F}_t = \{(\eta_s, \varepsilon_s') : s \le t\}$, and

$$\overline{w}_t(s_k) = w_t(s_k)\cdot \mathrm{I}\Big(\|\overline{v}_{t-1,*}(s_k)\| \le \frac{(nh)^{3/4}}{(\log n)^{1/4}}, |\eta_t| \le \frac{(nh)^{1/4}}{(\log n)^{1/4}}\Big),\quad \widetilde{w}_t(s_k) = w_t(s_k) - \overline{w}_t(s_k).$$

Note that

$$
\begin{aligned}
\|\mathbb{E}[\widetilde{w}_t(s_k)|\mathcal{F}_{t-1}]\| &= \left\|\mathbb{E}\left[\overline{v}_{t-1,*}(s_k)\eta_t \cdot \mathrm{I}\Big(\|\overline{v}_{t-1,*}(s_k)\| > \frac{(nh)^{3/4}}{(\log n)^{1/4}} \text{ or } |\eta_t| > \frac{(nh)^{1/4}}{(\log n)^{1/4}}\Big)\Big|\mathcal{F}_{t-1}\right]\right\| \\
&\le \left[\|\overline{v}_{t-1,*}(s_k)\|\,\mathrm{I}\Big(\|\overline{v}_{t-1,*}(s_k)\| > \frac{(nh)^{3/4}}{(\log n)^{1/4}}\Big)\right] \cdot \mathbb{E}\left[|\eta_t|\right] \\
&\quad + \|\overline{v}_{t-1,*}(s_k)\| \cdot \mathbb{E}\left[|\eta_t|\mathrm{I}\Big(|\eta_t| > \frac{(nh)^{1/4}}{(\log n)^{1/4}}\Big)\right] \\
&\equiv V_{t1}(s_k) + V_{t2}(s_k).
\end{aligned}
$$

As

$$
\mathbb{E}\left[|\eta_t|\,\mathrm{I}\Big(|\eta_t| > \frac{(nh)^{1/4}}{(\log n)^{1/4}}\Big)\right] \le \frac{(\log n)^{(3+\delta_0)/4}}{(nh)^{(3+\delta_0)/4}}\mathbb{E}\left[|\eta_t|^{4+\delta_0}\right]
$$

and $\|\overline{v}_{t-1,*}(s_k)\| = o_P\left((nh)^{3/4}\right)$ uniformly in $k = 1, \cdots, \overline{R}_n^*$ and $s_k - nh < t < s_k + nh$, we have

$$
\max_{1\le k\le \overline{R}_n^*}\left\|\sum_{t=1}^{n} V_{t2}(s_k)K\big(\frac{t-s_k}{nh}\big)\right\| = O_P\left(\frac{(nh)^{7/4}(\log n)^{(3+\delta_0)/4}}{(nh)^{(3+\delta_0)/4}}\right) = o_P\left(nh\sqrt{\log n}\right). \tag{A.29}
$$

On the other hand, as $\frac{n^{\delta_0-4}h^{\delta_0+12}}{(\log n)^{\delta_0}} \to \infty$, note that

$$
\begin{aligned}
& \mathbb{P}\left\{\max_{1\le k\le \overline{R}_n^*}\max_{s_k-nh\le t\le s_k+nh}\|\overline{v}_{t-1,*}(s_k)\|\mathrm{I}\Big(\|\overline{v}_{t-1,*}(s_k)\| > \frac{(nh)^{3/4}}{(\log n)^{1/4}}\Big) > 0\right\} \\
\le\ & \mathbb{P}\left\{\max_{1\le k\le \overline{R}_n^*}\max_{s_k-nh\le t\le s_k+nh}\|\overline{v}_{t-1,*}(s_k)\| > \frac{(nh)^{3/4}}{(\log n)^{1/4}}\right\} \\
\le\ & C \cdot \frac{n\overline{R}_n^*(\log n)^{(4+\delta_0)/4}}{(nh)^{(4+\delta_0)/4}} = o(1),
\end{aligned} \tag{A.30}
$$

and

$$
\begin{aligned}
& \mathbb{P}\left\{\max_{1\le k\le \overline{R}_n^*}\max_{s_k-nh\le t\le s_k+nh}\|\widetilde{w}_t(s_k)\| > 0\right\} \\
\le\ & \mathbb{P}\left\{\max_{1\le k\le \overline{R}_n^*}\max_{s_k-nh\le t\le s_k+nh}\|\overline{v}_{t-1,*}(s_k)\| > \frac{(nh)^{3/4}}{(\log n)^{1/4}}\right\} + \mathbb{P}\left\{\max_{1\le t\le n}|\eta_t| > \frac{(nh)^{1/4}}{(\log n)^{1/4}}\right\} \\
\le\ & C \cdot \frac{n\overline{R}_n^*(\log n)^{(4+\delta_0)/4}}{(nh)^{(4+\delta_0)/4}} + C \cdot \frac{n(\log n)^{(4+\delta_0)/4}}{(nh)^{(4+\delta_0)/4}} = o(1).
\end{aligned} \tag{A.31}
$$

Then, following the arguments in the proof of (A.12) and by (A.30) and (A.31), it follows that

$$
\max_{1\le k\le \overline{R}_n^*}\left\|\sum_{t=1}^{n} V_{t1}(s_k)K\big(\frac{t-s_k}{nh}\big)\right\| = o_P\left(nh\sqrt{\log n}\right) \tag{A.32}
$$

and

$$\max_{1\le k\le \overline{R}_n^*}\left\|\sum_{t=1}^{n}\widetilde{w}_t(s_k)K\big(\frac{t-s_k}{nh}\big)\right\| = o_P\left(nh\sqrt{\log n}\right). \tag{A.33}$$

Hence, by virtue of (A.29), (A.32) and (A.33),

$$\max_{1\le k\le \overline{R}_n^*}\left\|\sum_{t=1}^{n}\left(\widetilde{w}_t(s_k)-\mathbb{E}[\widetilde{w}_t(s_k)|\mathcal{F}_{t-1}]\right)K\big(\frac{t-s_k}{nh}\big)\right\| = o_P\left(nh\sqrt{\log n}\right). \tag{A.34}$$

On the other hand, note that $\{(w_t(s_k),\mathcal{F}_t):\ t\ge 1\}$ is a sequence of martingale differences. Let c_1 be some positive constant such that

$$\max_{1\le k\le \overline{R}_n^*}\left\|\mathrm{Var}\Big(\sum_{t=1}^{n}\left(\overline{w}_t(s_k)-\mathbb{E}[\overline{w}_t(s_k)|\mathcal{F}_{t-1}]\right)K\big(\frac{t-s_k}{nh}\big)\Big)\right\| \le c_1(nh)^2.$$

Then, by the exponential inequality for martingale differences (c.f., Theorem 1.2A in de la Peña, 1999), we have

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\le k\le \overline{R}_n^*}\left\|\sum_{t=1}^{n}\left(\overline{w}_t(s_k)-\mathbb{E}[\overline{w}_t(s_k)|\mathcal{F}_{t-1}]\right)K\big(\frac{t-s_k}{nh}\big)\right\| > M_1 nh\sqrt{\log n}\right\}\\
\le\ & \sum_{k=1}^{\overline{R}_n^*}\mathbb{P}\left\{\left\|\sum_{t=1}^{n}\left(\overline{w}_t(s_k)-\mathbb{E}[\overline{w}_t(s_k)|\mathcal{F}_{t-1}]\right)K\big(\frac{t-s_k}{nh}\big)\right\| > M_1 nh\sqrt{\log n}\right\}\\
\le\ & \sum_{k=1}^{\overline{R}_n^*}\exp\left\{-\frac{M_1^2(nh)^2\log n}{2[c_1(nh)^2+2c_K M_1(nh)^2]}\right\}\\
\le\ & O\left(\overline{r}_n^{-1}n^{-\sqrt{M_1}}\right) = o(1),
\end{aligned}$$

where M_1 is chosen such that

$$M_1^{3/2} > 2c_1 + 4c_K M_1 \ \text{ and }\ \overline{r}_n^{-1}n^{-\sqrt{M_1}} = o(1).$$

This indicates that

$$\max_{1\le k\le \overline{R}_n^*}\left\|\sum_{t=1}^{n}\left(\overline{w}_t(s_k)-\mathbb{E}[\overline{w}_t(s_k)|\mathcal{F}_{t-1}]\right)K\big(\frac{t-s_k}{nh}\big)\right\| = O_P\left(nh\sqrt{\log n}\right). \tag{A.35}$$

Then, by (A.34) and (A.35), result (A.28) follows.

For Ξ_{n1}, noting that

$$\sup_{s\in\mathbb{S}_k}\left|K\big(\frac{t-s}{nh}\big)-K\big(\frac{t-s_k}{nh}\big)\right| \le C\sup_{s\in\mathbb{S}_k}\left|\frac{s-s_k}{nh}\right| \le C\cdot\frac{n\overline{r}_n}{nh} = C\overline{r}_n h^{-1},$$

$\|\overline{v}_{t-1,*}(s)\| = o_P((nh)^{3/4})$ uniformly in $s \in \mathbb{S}_k$ for $k = 1, \cdots, \overline{R}_n^*$ and $s_k - nh < t < s + nh$, and $|\overline{e}_t| = O_P(1)$, we have

$$\begin{aligned}\Xi_{n1} &= \max_{1\le k\le \overline{R}_n^*} \sup_{s\in\mathbb{S}_k} \left\| \sum_{t=1}^n \overline{v}_{t-1,*}(s)\overline{e}_t \left[K\big(\frac{t-s}{nh}\big) - K\big(\frac{t-s_k}{nh}\big)\right]\right\| \\ &= o_P\left(n(nh)^{3/4}\overline{r}_n h^{-1}\right) = o_P\left(nh\sqrt{\log n}\right),\end{aligned}$$

which completes the proof of (A.26). For Ξ_{n2}, noting that $K(\cdot)$ is bounded, $|\overline{e}_t| = O_P(1)$ and

$$\max_{1\le k\le \overline{R}_n^*} \sup_{s\in\mathbb{S}_k} \|\overline{v}_{t-1,*}(s) - \overline{v}_{t-1,*}(s_k)\| = O_P\left(\sqrt{n\overline{r}_n}\right)$$

for any $s_k - nh < t < s + nh$, we then have

$$\begin{aligned}\Xi_{n2} &= \max_{1\le k\le \overline{R}_n^*} \sup_{s\in\mathbb{S}_k} \left\| \sum_{t=1}^n \left[\overline{v}_{t-1,*}(s) - \overline{v}_{t-1,*}(s_k)\right] \overline{e}_t K\big(\frac{t-s_k}{nh}\big)\right\| \\ &= O_P\left(n\sqrt{n\overline{r}_n}\right) = O_P\left(nh\sqrt{\log n}\right),\end{aligned} \tag{A.36}$$

which completes the proof of (A.27). Thus, (A.25) and then (A.16) have been proved. We have finally completed the proof of (A.5), and Theorem 2.1 then follows. □

Proof of Theorem 3.1. Note that

$$\begin{aligned}Q_n(z) &= \sum_{t=1}^n K\big(\frac{Z_t - z}{h}\big)X_{t-1}e_t + \sum_{t=1}^n K\big(\frac{Z_t - z}{h}\big)v_t e_t \\ &\equiv Q_{n1}(z) + Q_{n2}(z).\end{aligned} \tag{A.37}$$

First consider $Q_{n1}(z)$, which is the leading term of $Q_n(z)$. Decompose $Q_{n1}(z)$ as

$$\begin{aligned}Q_{n1}(z) &= \sum_{t=1}^n \mathbb{E}\big[K\big(\frac{Z_t - z}{h}\big)\big]X_{t-1}e_t + \sum_{t=1}^n \left\{K\big(\frac{Z_t - z}{h}\big) - \mathbb{E}\big[K\big(\frac{Z_t - z}{h}\big)\big]\right\} X_{t-1}e_t \\ &\equiv Q_{n3}(z) + Q_{n4}(z).\end{aligned} \tag{A.38}$$

Noting that by Assumptions 2 and 3,

$$\mathbb{E}\left[K\big(\frac{Z_t - z}{h}\big)\right] = \int K\big(\frac{z_1 - z}{h}\big) f_Z(z_1)dz_1 = h\int K(z_2) f_Z(z + z_2 h)dz_2 = hf_Z(z)\mu_0 + O(h^2),$$

uniformly for $0 \le z \le 1$, and by using the functional limit theorem for the partial sum of the linear process (Phillips and Solo, 1992) and Theorem 3.1 in Ibragimov and Phillips (2008), we have

$$\sum_{t=1}^n X_{t-1}e_t = O_P(n).$$

We deduce that

$$\sup_{0\le z\le 1} \big\|Q_{n3}(z)\big\| = O_P(nh) = o_P(n\sqrt{h\log n}). \tag{A.39}$$

For $Q_{n4}(z)$, it is easy to check that $\{(u_t(K,z)X_{t-1}e_t, \mathcal{F}_t^*)\}$ is a sequence of martingale differences, where

$$u_t(K,z) = K\big(\frac{Z_t - z}{h}\big) - \mathbb{E}\left[K\big(\frac{Z_t - z}{h}\big)\right], \quad \mathcal{F}_t^* = \sigma\left\{\eta_{t+1}, (Z_s, \eta_s, \varepsilon_s) : s \le t\right\}.$$

The following proof is similar to the proof of (A.25) with some modifications. We cover the interval $[0,1]$ by a finite number of disjoint intervals $\mathbb{Z}_k$ with centre point z_k and radius r_n defined in the proof of Theorem 2.1, and the number of these intervals is $N_n = O(r_n^{-1})$. By some standard arguments, we have

$$\begin{aligned}\sup_{0\le z\le 1}\left\|\sum_{t=1}^n u_t(K,z)X_{t-1}e_t\right\| \quad &\le \quad \max_{1\le k\le N_n}\sup_{z\in\mathbb{Z}_k}\left\|\sum_{t=1}^n X_{t-1}e_t\left[u_t(K,z) - u_t(K,z_k)\right]\right\| + \\ &\quad\quad \max_{1\le k\le N_n}\left\|\sum_{t=1}^n u_t(K,z_k)X_{t-1}e_t\right\|.\end{aligned}$$

Noting that by Assumption 2(i)

$$|u_t(K,z) - u_t(K,z_k)| \le 2\left|K\big(\frac{Z_t - z}{h}\big) - K\big(\frac{Z_t - z_k}{h}\big)\right| = O_P\left(r_n h^{-1}\right),$$

and $\max_{1\le t\le n}\|X_t\| = O_P(\sqrt{n})$, we have

$$\max_{1\le k\le N_n}\sup_{z\in\mathbb{Z}_k}\left\|\sum_{t=1}^n X_{t-1}e_t\left[u_t(K,z) - u_t(K,z_k)\right]\right\| = O_P\left(n^{3/2}r_n h^{-1}\right) = O_P\left(n\sqrt{h\log n}\right). \quad \text{(A.40)}$$

We next prove that

$$\max_{1\le k\le N_n}\left\|\sum_{t=1}^n u_t(K,z_k)X_{t-1}e_t\right\| = O_P\left(n\sqrt{h\log n}\right). \quad \text{(A.41)}$$

As $\frac{n^{2+\delta_0}h^{4+\delta_0}}{(\log n)^{4+\delta_0}} \to \infty$, there exists a positive function $l(n)$ such that

$$l(n) \to \infty \quad \text{and} \quad \frac{n^{2+\delta_0}h^{4+\delta_0}}{l(n)(\log n)^{4+\delta_0}} \to \infty. \quad \text{(A.42)}$$

Let $W_t(z_k) = u_t(K,z_k)X_{t-1}e_t$, $L(n) = \left[l(n)\right]^{\frac{1}{4+\delta_0}}$, and

$$\overline{W}_t(z_k) = W_t(z_k)\cdot \mathrm{I}\left(\|X_{t-1}\| \le \sqrt{nL(n)}, |e_t| \le \sqrt{\frac{nh}{L(n)\log n}}\right), \quad \widetilde{W}_t(z_k) = W_t(z_k) - \overline{W}_t(z_k).$$

From the definition of $\widetilde{W}_t(z_k)$, it is easy to see that if the two events $\left\{\|X_{t-1}\| \le \sqrt{nL(n)}, t = 1,\cdots,n\right\}$ and $\left\{|e_t| \le \sqrt{\frac{nh}{L(n)\log n}}, t = 1,\cdots,n\right\}$ hold simultaneously, $\left\|\sum_{t=1}^n \widetilde{W}_t(z_k)\right\| = 0$ for any $1 \le k \le N_n$.

In other words, if $\left\|\sum_{t=1}^n \widetilde{W}_t(z_k)\right\| > 0$, we must have either $\left\{\|X_{t-1}\| > \sqrt{nL(n)}\right\}$ for at least one $1 \le t \le n$, or $\left\{|e_t| > \sqrt{\frac{nh}{L(n)\log n}}\right\}$ for at least one $1 \le t \le n$. Hence, we have for any $\epsilon > 0$,

$$\begin{aligned}
& \mathbb{P}\left\{\max_{1\le k\le N_n}\left\|\sum_{t=1}^n \widetilde{W}_t(z_k)\right\| > \epsilon n\sqrt{h\log n}\right\} \\
\le\ & \mathbb{P}\left\{\max_{1\le t\le n}\|X_{t-1}\| > \sqrt{nL(n)}\right\} + \mathbb{P}\left\{\max_{1\le t\le n}|e_t| > \sqrt{\frac{nh}{L(n)\log n}}\right\} \\
=\ & o(1) + O\left(\frac{n[L(n)\log n]^{(4+\delta_0)/2}}{(nh)^{(4+\delta_0)/2}}\right) = o(1), \qquad \text{(A.43)}
\end{aligned}$$

by (A.42), and (A.43) leads to

$$\max_{1\le k\le N_n}\left\|\sum_{t=1}^n \widetilde{W}_t(z_k)\right\| = o_P\left(n\sqrt{h\log n}\right). \qquad \text{(A.44)}$$

Let c_2 be some positive constant such that

$$\max_{1\le k\le N_n}\left\|\mathrm{Var}\big(\sum_{t=1}^n \overline{W}_t(z_k)\big)\right\| \le c_2 n^2 h.$$

Then, by the exponential inequality for martingale differences again (c.f., Theorem 1.2A in de la Peña, 1999), we have

$$\begin{aligned}
& \mathbb{P}\left\{\max_{1\le k\le N_n}\left\|\sum_{t=1}^n \overline{W}_t(z_k)\right\| > M_2 n\sqrt{h\log n}\right\} \le \sum_{k=1}^{N_n}\mathbb{P}\left\{\left\|\sum_{t=1}^n \overline{W}_t(z_k)\right\| > M_2 n\sqrt{h\log n}\right\} \\
\le\ & \sum_{k=1}^{N_n}\exp\left\{-\frac{M_2^2 n^2 h\log n}{2(c_2 n^2 h + 2c_K M_2 n^2 h)}\right\} = O\left(r_n^{-1} n^{-\sqrt{M_2}}\right) = o(1),
\end{aligned}$$

where M_2 is chosen such that

$$M_2^{3/2} > 2c_2 + 4c_K M_2, \quad r_n^{-1} n^{-\sqrt{M_2}} = o(1).$$

This indicates that

$$\max_{1\le k\le N_n}\left\|\sum_{t=1}^n \overline{W}_t(z_k)\right\| = O_P\left(n\sqrt{h\log n}\right). \qquad \text{(A.45)}$$

In view of (A.44) and (A.45), we can complete the proof of (A.41), which together with (A.40), indicates that

$$\sup_{0\le z\le 1}\|Q_{n4}(z)\| = O_P\left(n\sqrt{h\log n}\right). \qquad \text{(A.46)}$$

Then, by (A.39) and (A.46), we deduce that

$$\sup_{0\le z\le 1}\|Q_{n1}(z)\| = \sup_{0\le z\le 1}\|Q_{n3}(z)\| + \sup_{0\le z\le 1}\|Q_{n4}(z)\| = O_P\big(n\sqrt{h\log n}\big). \qquad \text{(A.47)}$$

We next consider $Q_{n2}(z)$, which is relatively simpler. Let $v_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \varepsilon_{t-j} = \varepsilon_t + \sum_{j=1}^{\infty} \mathbf{\Phi}_j \varepsilon_{t-j} \equiv \varepsilon_t + \widehat{v}_t$ and $e_t = \sum_{j=0}^{\infty} \phi_j \eta_{t-j} = \eta_t + \sum_{j=1}^{\infty} \phi_j \eta_{t-j} \equiv \eta_t + \widehat{e}_t$. Note that

$$\begin{aligned} Q_{n2}(z) &= \sum_{t=1}^{n} K\big(\frac{Z_t - z}{h}\big)\varepsilon_t \eta_t + \sum_{t=1}^{n} K\big(\frac{Z_t - z}{h}\big)\widehat{v}_t \eta_t + \\ & \quad \sum_{t=1}^{n} K\big(\frac{Z_t - z}{h}\big)\varepsilon_t \widehat{e}_t + \sum_{t=1}^{n} K\big(\frac{Z_t - z}{h}\big)\widehat{v}_t \widehat{e}_t \\ &\equiv \sum_{k=5}^{8} Q_{nk}(z). \end{aligned} \tag{A.48}$$

We next prove that

$$\sup_{0 \le z \le 1} \|Q_{nk}(z)\| = o_P\left(n\sqrt{h \log n}\right), \quad k = 5, 6, 7, 8. \tag{A.49}$$

For $k = 8$, note the decompositions:

$$Q_{n8}(z) = \sum_{t=1}^{n} \mathbb{E}\left[K\big(\frac{Z_t - z}{h}\big)\right]\widehat{v}_t \widehat{e}_t + \sum_{t=1}^{n} \left\{K\big(\frac{Z_t - z}{h}\big) - \mathbb{E}\left[K\big(\frac{Z_t - z}{h}\big)\right]\right\}\widehat{v}_t \widehat{e}_t.$$

By the fact that $\mathbb{E}\left[K\big(\frac{Z_t - z}{h}\big)\right] = h f_Z(z)\mu_0 + O(h^2)$ uniformly in $0 \le z \le 1$ and $\{\widehat{v}_t \widehat{e}_t\}$ is a sequence of stationary random vectors, and using the ergodic theorem, we have

$$\sup_{0 \le z \le 1} \left\|\sum_{t=1}^{n} \mathbb{E}\left[K\big(\frac{Z_t - z}{h}\big)\right]\widehat{v}_t \widehat{e}_t\right\| = O_P(nh) = o_P\left(n\sqrt{h \log n}\right). \tag{A.50}$$

On the other hand, note that $\{(u_t(K, z)\widehat{v}_t \widehat{e}_t, \mathcal{F}_t^*)\}$ is a sequence of martingale differences, where $u_t(K, z) = K\big(\frac{Z_t - z}{h}\big) - \mathbb{E}\left[K\big(\frac{Z_t - z}{h}\big)\right]$ and $\mathcal{F}_t^* = \sigma\{\eta_{t+1}, (Z_s, \eta_s, \varepsilon_s) : s \le t\}$ are defined as above. Then following the argument in the proof of (A.46) with some modification, we can prove that

$$\sup_{0 \le z \le 1} \left\|\sum_{t=1}^{n} \left\{K\big(\frac{Z_t - z}{h}\big) - \mathbb{E}\left[K\big(\frac{Z_t - z}{h}\big)\right]\right\}\widehat{v}_t \widehat{e}_t\right\| = o_P\left(n\sqrt{h \log n}\right). \tag{A.51}$$

By (A.50) and (A.51), we have completed the proof of (A.49) with $k = 8$. The proof for other cases of $k = 5, 6, 7$ is analogous and thus the details are omitted here. Then, by (A.49), we have

$$\sup_{0 \le z \le 1} \|Q_{n2}(z)\| \le \sum_{k=5}^{8} \sup_{0 \le z \le 1} \|Q_{nk}(z)\| = o_P\left(n\sqrt{h \log n}\right). \tag{A.52}$$

By (A.37), (A.47) and (A.52), we can prove (3.2). Then, the proof of Theorem 3.1 can be completed.
$\square$

Proof of Theorem 4.1. Note that

$$
\begin{aligned}
\widehat{\beta}_n(z) - \beta(z) &= \left[\sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big)\right]^{+} \left\{\sum_{t=1}^{n} X_t X_t' \left[\beta\big(\frac{t}{n}\big) - \beta(z)\right] K\big(\frac{t-nz}{nh}\big)\right\} + \\
&\quad \left[\sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big)\right]^{+} \left[\sum_{t=1}^{n} X_t e_t K\big(\frac{t-nz}{nh}\big)\right] \\
&\equiv \Pi_{n1}(z) + \Pi_{n2}(z).
\end{aligned} \tag{A.53}
$$

We first prove that the matrix

$$
\mathbf{R}_n^{-1}\mathbf{D}_n(z)' \left[\sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big)\right] \mathbf{D}_n(z)\mathbf{R}_n^{-1}
$$

is asymptotically non-singular for all $z \in [\epsilon_\diamond, 1-\epsilon_\diamond]$, where $\epsilon_\diamond$ is defined in Assumption 5. This can be proved by combining Proposition A.1 and Lemma B.4 in Phillips *et al* (2013). In order to keep the paper self-contained, we outline the proof. Recall that $q_z = q_{\gamma_n(z)}$ and $q_z^\perp = q_{\gamma_n(z)}^\perp$, and observe that

$$
\begin{aligned}
& \mathbf{R}_n^{-1}\mathbf{D}_n(z)' \left[\sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big)\right] \mathbf{D}_n(z)\mathbf{R}_n^{-1} \\
= & \begin{bmatrix} q_z' \left[\frac{1}{n^2 h} \sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big)\right] q_z & q_z' \left[\frac{1}{n^2 h^{3/2}} \sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big)\right] q_z^\perp \\ (q_z^\perp)' \left[\frac{1}{n^2 h^{3/2}} \sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big)\right] q_z & (q_z^\perp)' \left[\frac{1}{n^2 h^{2}} \sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big)\right] q_z^\perp \end{bmatrix} \\
\equiv & \begin{bmatrix} \Delta_{n1}(z) & \Delta_{n2}(z) \\ \Delta_{n2}(z)' & \Delta_{n3}(z) \end{bmatrix}.
\end{aligned} \tag{A.54}
$$

For $\Delta_{n1}(z)$, following the argument in the proof of Proposition A.1 in Phillips *et al* (2013) and using the definitions of $b_{\gamma_n(z)}$ and q_z, we can claim that

$$
\begin{aligned}
\frac{1}{n^2 h}\sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big) &= \left(\frac{1}{\sqrt{n}} X_{\gamma_n(z)}\right)\left(\frac{1}{\sqrt{n}} X_{\gamma_n(z)'}\right)\left[\frac{1}{nh}\sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)\right] + o_P(1) \\
&= b_{\gamma_n(z)} b_{\gamma_n(z)}' + o_P(1) = \|b_{\gamma_n(z)}\|^2 q_z q_z' + o_P(1)
\end{aligned} \tag{A.55}
$$

uniformly in z. Then, by Assumption 1(i), the BN decomposition, and the strong approximation result (e.g. Csörgö and Révész, 1981), there exists $B_z(\mathbf{\Omega}_\varepsilon)$ such that

$$
\sup_{\epsilon_\diamond \le z \le \epsilon_\diamond} \|q_z - \overline{q}_z\| + \sup_{\epsilon_\diamond \le z \le \epsilon_\diamond} \left\|b_z - \overline{b}_z\right\| = o_P(1), \tag{A.56}
$$

where $\bar{b}_z = B_z(\mathbf{\Omega}_\varepsilon)$ and $\bar{q}_z = \bar{b}_z/\|\bar{b}_z\|$. Then, by (A.55), (A.56) and the definition of $\Delta_{n1}(z)$, we can prove that

$$\sup_{\epsilon_\diamond \le z \le \epsilon_\diamond} |\Delta_{n1}(z) - \Delta_1(z)| = o_P(1), \tag{A.57}$$

where $\Delta_1(z)$ is defined in Section 4.

For $\Delta_{n2}(z)$, following the argument in the proof of Proposition A.1 in Phillips *et al* (2013) again and using the orthogonality condition (2.4), the asymptotic leading term of $\Delta_{n2}(z)$ is:

$$q_z' q_z \left[\frac{1}{n^{3/2}h^{3/2}} \sum_{t=1}^{n} \overline{v}_t(z)' K\big(\frac{t-nz}{nh}\big) \right] q_z^\perp, \tag{A.58}$$

where $\overline{v}_t(z)$ is defined in the proof of (A.5) and independent of q_z and $q_z^\perp$. Then, by Assumption 1(i) and the strong approximation result again, there exists $B_{z,*}(\mathbf{\Omega}_\varepsilon)$ (which is also independent of q_z and $q_z^\perp$) such that

$$\sup_{\epsilon_\diamond \le z \le \epsilon_\diamond} \sup_{\gamma_n(z) \le t \le \gamma_n(z)+2nh} \left\| \frac{1}{\sqrt{2nh}} \overline{v}_t(z) - B_{m(t,z),*}(\mathbf{\Omega}_\varepsilon) \right\| = o_P(1), \tag{A.59}$$

where $m(t,z) = \frac{t-\gamma_n(z)}{2nh}$. Then, by (A.56), (A.58) and (A.59), we have

$$\sup_{\epsilon_\diamond \le z \le \epsilon_\diamond} |\Delta_{n2}(z) - \Delta_2(z)| = o_P(1), \tag{A.60}$$

where $\Delta_2(z)$ is defined in Section 4. By an analogous argument and using the orthogonality condition (2.4), we can also show that

$$\sup_{\epsilon_\diamond \le z \le \epsilon_\diamond} |\Delta_{n3}(z) - \Delta_3(z)| = o_P(1), \tag{A.61}$$

where $\Delta_3(z)$ is defined in Section 4. Then, by the asymptotic non-singularity of $\mathbf{\Delta}_z$ in Assumption 5, (A.57), (A.60) and (A.61), we can prove that $\mathbf{R}_n^{-1}\mathbf{D}_n(z)' \left[\sum_{t=1}^n X_t X_t' K\big(\frac{t-nz}{nh}\big)\right] \mathbf{D}_n(z)\mathbf{R}_n^{-1}$ is asymptotically non-singular for all $\epsilon_\diamond \le z \le \epsilon_\diamond$, which together with Theorem 2.1, implies that

$$\sup_{\epsilon_\diamond \le z \le \epsilon_\diamond} \left\| \mathbf{R}_n \mathbf{D}_n(z)' \Pi_{n2}(z) \right\| = O_P(\sqrt{\log n}). \tag{A.62}$$

The above convergence result leads to

$$\sup_{\epsilon_\diamond \le z \le \epsilon_\diamond} \left| q_z' \Pi_{n2}(z) \right| = O_P\left(\sqrt{\frac{\log n}{n^2 h}} \right), \quad \sup_{\epsilon_\diamond \le z \le \epsilon_\diamond} \left\| (q_z^\perp)' \Pi_{n2}(z) \right\| = O_P\left(\frac{\sqrt{\log n}}{nh} \right). \tag{A.63}$$

By Assumption 4, we can show that

$$\left\| \beta\big(\frac{t}{n}\big) - \beta(z) \right\| = O(h^{\alpha_0}), \quad \left| \frac{t}{n} - z \right| \le h. \tag{A.64}$$

Following the argument above and using (A.64), we can prove that the asymptotic order of

$$\mathbf{R}_n^{-1}\mathbf{D}_n(z)'\left\{\sum_{t=1}^{n} X_tX_t'\left[\beta(\frac{t}{n}) - \beta(z)\right] K\big(\frac{t-nz}{nh}\big)\right\}\mathbf{D}_n(z)\mathbf{R}_n^{-1}$$

is $O_P(h^{\alpha_0})$, which together with the fact that $\mathbf{R}_n^{-1}\mathbf{D}_n(z)'\left[\sum_{t=1}^{n} X_tX_t'K\big(\frac{t-nz}{nh}\big)\right]\mathbf{D}_n(z)\mathbf{R}_n^{-1}$ is asymptotically non-singular, implies that

$$\sup_{\epsilon_\diamond\le z\le\epsilon_\diamond} \|\Pi_{n1}(z)\| = O_P(h^{\alpha_0}). \tag{A.65}$$

The proof of Theorem 4.1 can be completed in view of (A.53), (A.63), and (A.65) in conjunction with the definitions of $\mathbf{R}_n$ and $\mathbf{D}_n(z)$. □

Proof of Theorem 4.2. The proof is similar to the proof of Theorem 4.1 above. As in (A.45), we have

$$\begin{aligned}\widehat{\beta}_n(z) - \beta(z) &= \left[\sum_{t=1}^{n} X_tX_t'K\big(\frac{Z_t-z}{h}\big)\right]^{+}\left\{\sum_{t=1}^{n} X_tX_t'\left[\beta(Z_t) - \beta(z)\right]K\big(\frac{Z_t-z}{h}\big)\right\} + \\ &\quad \left[\sum_{t=1}^{n} X_tX_t'K\big(\frac{Z_t-z}{h}\big)\right]^{+}\left[\sum_{t=1}^{n} X_te_tK\big(\frac{Z_t-z}{h}\big)\right] \\ &\equiv \Pi_{n3}(z) + \Pi_{n4}(z).\end{aligned} \tag{A.66}$$

Following the proof of Proposition A.1 in Li *et al* (2014), we can show that the random denominator $\frac{1}{n^2h}\sum_{t=1}^{n} X_tX_t'K\big(\frac{Z_t-z}{h}\big)$ is non-singular with probability 1 for all $z\in[0,1]$. We next give an outline of this proof. Note that

$$\begin{aligned}\frac{1}{n^2h}\sum_{t=1}^{n} X_tX_t'K\big(\frac{Z_t-z}{h}\big) &= \frac{1}{n^2h}\sum_{t=1}^{n} X_tX_t'\left\{K\big(\frac{Z_t-z}{h}\big) - \mathbb{E}\left[K\big(\frac{Z_t-z}{h}\big)\right]\right\} + \\ &\quad \frac{1}{n^2h}\sum_{t=1}^{n} X_tX_t'\mathbb{E}\left[K\big(\frac{Z_t-z}{h}\big)\right] \\ &\equiv \Pi_{n5}(z) + \Pi_{n6}(z).\end{aligned} \tag{A.67}$$

By the fact that $\mathbb{E}\left[K\big(\frac{Z_t-z}{h}\big)\right] = hf_Z(z)\mu_0 + o(h^2)$ and using continuous mapping, it follows that uniformly for $z\in[0,1]$,

$$\Pi_{n6}(z) = \frac{f_Z(z)}{n}\sum_{t=1}^{n}\frac{X_t}{\sqrt{n}}\cdot\frac{X_t'}{\sqrt{n}} + O_P(h) \Rightarrow f_Z(z)\int_0^1 B_r(\mathbf{\Omega}_\varepsilon)B_r(\mathbf{\Omega}_\varepsilon)'dr. \tag{A.68}$$

On the other hand, following the technical argument in Li *et al* (2014), we can prove that

$$\sup_{0\le z\le 1} \|\Pi_{n5}(z)\| = o_P(1), \tag{A.69}$$

which together with (A.67) and (A.68), implies that

$$\frac{1}{n^2h}\sum_{t=1}^{n} X_t X_t' K\Big(\frac{Z_t-z}{h}\Big) \Rightarrow f_Z(z)\int_0^1 B_r(\boldsymbol{\Omega}_\varepsilon)B_r(\boldsymbol{\Omega}_\varepsilon)'dr \tag{A.70}$$

uniformly for $z \in [0,1]$. As $\inf_{0\le z\le 1} f_Z(z) > 0$ and using the fact that $\int_0^1 B_r(\boldsymbol{\Omega}_\varepsilon)B_r(\boldsymbol{\Omega}_\varepsilon)'dr$ is positive definite with probability 1 by Lemma A2 in Phillips and Hansen (1990), we can claim that $\frac{1}{n^2h}\sum_{t=1}^{n} X_t X_t' K\big(\frac{Z_t-z}{h}\big)$ is asymptotically non-singular over $z \in [0,1]$, which together with Theorem 3.1, implies that

$$\sup_{0\le z\le 1}\|\Pi_{n4}(z)\| = O_P\left(\sqrt{\frac{\log n}{n^2h}}\right). \tag{A.71}$$

On the other hand, by Assumption 4 and following the proof of (A.65), it follows easily that

$$\sup_{0\le z\le 1}\|\Pi_{n3}(z)\| = O_P\left(h^{\alpha_0}\right). \tag{A.72}$$

Then, (4.7) can be proved by using (A.66), (A.71) and (A.72) and the proof of Theorem 4.2 is thus completed. $\square$